REMARKS

Claims 49, 54, 56-58, 63, 66, 72, 75, 77, and 79-80 are currently pending. Claims 56, 58 and 75 are amended herein. Claim 77 is canceled. Support for the claim amendments is provided by the specification at, for example, page 48, lines 14-20, and page 122, Table I. No new matter is added by way of the claim amendments. Applicants respectfully request entry of the claim amendments and reconsideration in view of the following. After entry of the claim amendments, claims 49, 54, 56-58, 63, 66, 72, 75, and 79-80 will be pending.

Rejection under 35 U.S.C. § 112, first paragraph – Enablement

Claims 49, 54-63, 66, 72, 75, 77, 79, and 80 remain rejected under 35 U.S.C. § 112, first paragraph, as allegedly lacking enablement. Applicants note that claims 55 and 59-62 were previously canceled. Applicants respectfully traverse the rejections with respect to the remaining claims.

As a preliminary matter, "[t]o be enabling, the specification of a patent must teach those skilled in the art to make and use the full scope of the claimed invention without 'undue experimentation' ... Nothing more than objective enablement is required, and therefore it is irrelevant whether this teaching is provided through broad terminology or illustrative examples." *In re Wright*, 999 F.2d 1557, 1561 (Fed. Cir. 1993). "The test is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed." *In re Wands*, 858 F.2d 731, 737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988) (citing *In re Angstadt*, 537 F.2d 489, 502-04, 190 USPQ 214, 217-19 (CCPA 1976)).

Claims to polypeptides of SEQ ID NOs:3, 5, and 7, peptides, and methods of detecting cancer

In the first grounds of rejection, claims 49, 54, 56, 57, 75 and 77, drawn to polypeptides of SEQ ID NOs:3, 5, and 7, peptides, and methods of detecting cancer are rejected as allegedly lacking enablement. Specifically, the Office has alleged that the specification fails to provide a nexus

between the expression of 254P1D6B in cancerous tissues and the expression of the individual variant sequences claimed, and that undue experimentation would be required to determine the expression pattern of the claimed variants before the variant clones could be used as markers for cancer. The Office further asserts an even greater degree of undue experimentation would be required to determine if the claimed sequences are over-expressed in cancers other than those listed on page 122 in order to practice the full scope of the claimed method. Applicants respectfully disagree.

Solely to advance prosecution, and without acquiescing to the Examiner's arguments, claim 75 is amended to claim a method for detecting the presence of cancer in the tissues listed in Table I. *See* specification at page 122, Table I. Claim 77 is canceled.

As previously noted, the specification provides Northern blot and PCR data for the so-called generic 254P1D6B protein (SEQ ID NO:3), which demonstrate that the gene of interest is upregulated in a variety of cancerous tissues relative to normal tissues. *See* the specification at, e.g., page 82, Example 4, and Figures 14 to 16. Applicants respectfully note that the use of Northern hybridization to measure gene expression was routine in the art at the time of the application. *See*, e.g., Sambrook & Russell, MOLECULAR CLONING: A LABORATORY MANUAL, 3rd ED. (2001), Cold Spring Harbor Laboratory Press, at 7.21, 2nd paragraph (noting that "Northern hybridization became part of the standard repertoire of molecular biology almost immediately after the first descriptions of the method were published.") (attached as Exhibit A).

Based on the expression data provided for the polypeptide of SEQ ID NO:3, the guidance provided by the specification, and the state of the art at the time of the application, applicants respectfully submit that the generation of expression data for the claimed polypeptides does not constitute "undue" experimentation.

The Examiner has also asserted that there is no objective evidence that all the variants possess the same properties as the generic 254P1D6B sequence (SEQ ID NO:3), or that the generic sequence is over-expressed in malignancies such as sarcomas, melanomas, etc.

Applicants respectfully note that genetic variation between the sequences does not necessarily imply that each variant has a unique function. Moreover, the biological function of the individual protein variants is irrelevant to the use of the claimed proteins as a family of markers for the detection of cancer by monitoring expression levels in test tissues. The latter issue, regarding over-expression in malignancies such as sarcomas, etc., is obviated by the amendment to claim 75.

Finally, the Examiner asserts that the specification does not teach how to use the peptides of claims 54 and 56 if said peptides do not generate an antibody which binds a polypeptide associated with a cancerous state. (*See* Office action at page 3).

Applicants respectfully note that claims 54 and 56 do not require that the claimed peptides generate an antibody which binds a polypeptide associated with a cancerous state. Rather, claims 54 and 56 require that the claimed peptides induce a specific antibody response against a polypeptide having the amino acid sequence of SEQ ID NO:3, 5, or 7.

In view of the foregoing, applicants respectfully submit that the claims, as amended, are fully enabled. Applicants respectfully request that the enablement rejection as it relates to claims 49, 54, 56, 57 and 75 be withdrawn. Claim 77 has been canceled, rendering the rejection moot.

Claims to methods of generating an immune response

Under the second grounds for rejection, the Office action states that claims 58-62, 79 and 80 are drawn to a method of generating an immune response in a mammal comprising exposing cells of said mammal to a polynucleotide of SEQ ID NO:3, 5, or 7. (*See* Office action at page 4).

Applicants respectfully note that claims 59-62 were previously canceled. Claims 79 and 80 are drawn to polynucleotides encoding the polypeptides of SEQ ID NOs:3, 5, and 7, wherein the polypeptide is encoded as a portion of a viral vector. Clarification of the grounds of rejection with respect to claims 79 and 80 is requested. Applicants will address the third grounds of rejection as it relates to claim 58.

The Examiner asserts that the claims encompass the generation of cytotoxic T cells which kill the autologous cells which express said proteins. The Examiner also asserts that it is well known in the art that an antibody must bind to a cell surface target, and cites Abbas et al. to support the contention that targets which evoke complement or ADCC-mediate cell killing must be on the cell surface. The Examiner further asserts that there is no evidence in the specification that the 254P1D6B protein is a cell surface protein, and that undue experimentation, without a reasonable expectation of success, would be required for one of skill in the art to use the claimed methods. Applicants respectfully disagree.

Without acquiescing the Examiner's argument, claim 58 is amended to clarify that the claim relates to a method of generating an immune response to a polypeptide having SEQ ID NO: 3, 5, or 7, wherein said immune response is the activation of B cells.

For reasons of record, applicants respectfully submit that the Examiner's argument mischaracterizes Abbas et al. Moreover, the specification provides sufficient evidence such that a person of skill in the art would reasonably conclude that the 254P1D6B protein is present on the cell surface. Secondary structure and transmembrane (TM) domain predictions for the protein of 254P1D6B v.1 (SEQ ID NO:3) are provided in Figure 13. For example, Figure 13C provides a schematic representation of the probability of the existence of a transmembrane region in the polypeptide of SEQ ID NO:3, based on the TMHMM algorithm of Sonnhammer et al., indicating that the polypeptide of SEQ ID NO:3 contains a single TM domain. *See* specification at, for example, page 7 and Figure 13C.

At the time of the application, the use of computational methods to predict protein secondary structure and the presence of transmembrane domains was well-known in the art. *See* Chen et al., *Protein Sci.* 2002; 11:2774-2791 (attached as Exhibit B). Chen et al. reported that the TMHMM algorithm correctly identified ca. 90% of all observed membrane helices. *See* Chen et al., at page 2777, left column and Table 1, and page 2778, Table 2 ($Q_{htm}^{\%obs}$, % of all observed helices that are predicted correctly). Based on the data provided in the specification for the polypeptide of SEQ ID NO:3, and the relatively high degree of predictive accuracy for the computational methods

described, a person of skill in the art would have a reasonable expectation of success that the claimed polypeptides would be detectable using an antibody exposed to such a cell.

The Examiner has cited copious references in support of the assertion that the state of the art with respect to treating patients with cancer by means of administering tumor antigens is unpredictable. Applicants respectfully note that the pending claims do not relate to methods of treating cancer patients by administering tumor antigens. The Examiner has further asserted that the specification does not provide any disclosure that the administration of the claimed polypeptides would generate CTLs. As amended, claim 58 recites a method of generating an immune response wherein said immune response is the activation of B cells. Accordingly, this basis of rejection may be properly withdrawn.

Claims to polynucleotides and host cells

Claims 63, 66, and 72, drawn to polynucleotides that encode polypeptides of SEQ ID NOs:3, 5, and 7, polynucleotides having SEQ ID NOs:2, 4, and 6, and host cells, are rejected for the reasons stated by the Office with respect to the polypeptide claims. Specifically, the Office alleges that no specific data has been presented for the individual polynucleotides of SEQ ID NOs:2, 4, and 6 to provide a nexus between the claimed polynucleotides and the detection of cancer. Applicants respectfully disagree, for the reasons stated above with respect to the claimed polypeptides.

As noted above, the specification provides actual data using sequences falling within the scope of the rejected claims to detect the over-expression of the gene of interest in cancer samples versus normal samples. *See* the specification at, e.g., page 82, Example 4, and Figures 14 to 16. The data provided by the applicants clearly demonstrates how one of ordinary skill in the art could use the claimed polynucleotides to detect cancer in a test tissue sample.

In view of the data and guidance provided by the specification, as well as the routine nature of such experiments, applicants respectfully submit that the experimentation required to generate data for the individual claimed polynucleotide sequences is not "undue". Accordingly, Applicants request that the enablement rejection as it relates to claims 63, 66, and 72 be withdrawn.

Rejection under 35 U.S.C. § 112, first paragraph – Written Description

Claims 54 and 56 stand rejected under 35 U.S.C. § 112, first paragraph as allegedly failing to comply with the written description requirement. Applicants note that claim 55 has been previously canceled. Applicants respectfully traverse the rejection.

To satisfy the written description requirement, a patent application must describe the invention in sufficient detail that one of skill in the relevant art could reasonably conclude that the inventor was in possession of the claimed invention at the time the application was filed. *See Vas-Cath Inc. v. Mahurkar*, 935 F.2d 1555, 1563-64, (Fed. Cir. 1991). An applicant need not describe exactly the subject matter claimed in the specification in order to satisfy the written description requirement. *See Union Oil of Cal. v. Atlantic Richfield Co.*, 208 F.3d 989, 997 (Fed. Cir. 2000). "What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail." See *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94.

In the applicants' response submitted on January 30, 2007, the applicants noted that the specification described polypeptide sequences which could be used to produce the claimed peptides. *See* specification at, for example, Figure 2. The applicants also described how to use the claimed peptides to induce a specific antibody response, as well as how to test for such a response. *See* specification at, for example, pages 90-92, Examples 10 and 11. Applicants have also identified specific amino acid regions of 254P1D6B variant 1 (SEQ ID NO:3) which may be used to induce an antibody response. *See* specification at, for example, page 90, Example 10, para. 2.

Notwithstanding the foregoing, the Examiner has stated that because the polypeptides of SEQ ID NOs:3, 5, and 7, and the claimed peptide fragments were not *themselves* conventionally known in the art, that a connection between the epitopes and the induction of a specific antibody response cannot be inferred. (*See* Office action at page 9). Applicants respectfully disagree.

The induction of specific antibody responses to peptides was conventionally known in the art at the time the present application was filed. *See* specification at, for example, page 63, line 36 to page 64, line 4; *see also* Ausubel, et al., Eds., *Current Protocols In Molecular Biology* (2002),

John Wiley & Sons, Inc., at VOL. 2, 11.16.1-11.16.5, and 11.16.16-11.16.19 (attached as Exhibit C). What has not been literally described by the applicants was thus conventionally known in the art at the time the present application was filed.

In view of the applicant's disclosure, as well as what was known in the art at the time of the application, a person of ordinary skill in the art could readily conclude that applicants were in possession of the claimed invention when the application was filed. Accordingly, applicants respectfully request that the written description rejection under 35 U.S.C. § 112, first paragraph, be withdrawn.

CONCLUSION

In view of the above, each of the presently pending claims in this application is believed to be in immediate condition for allowance. Accordingly, the Examiner is respectfully requested to withdraw the outstanding rejection of the claims and to pass this application to issue. If it is determined that a telephone conference would expedite the prosecution of this application, the Examiner is invited to telephone the undersigned at the number given below.

In the event the U.S. Patent and Trademark office determines that an extension and/or other relief is required, applicants petition for any required relief including extensions of time and authorize the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to **Deposit Account No. 03-1952** referencing docket no. **511582008100**. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: June 11, 2007

Respectfully submitted,

By: Elizabeth Cary Miller

Elizabeth Cary Miller
Registration No.: 54,708
MORRISON & FOERSTER LLP
12531 High Bluff Drive
Suite 100
San Diego, California 92130-2040
(858) 314-7717

VOLUME 1

Molecular Cloning

A LABORATORY MANUAL

THIRD EDITION

www.MolecularCloning.com

Joseph Sambrook
PETER MACCALLUM CANCER INSTITUTE AND THE UNIVERSITY OF MELBOURNE, AUSTRALIA

David W. Russell
UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER, DALLAS

COLD SPRING HARBOR LABORATORY PRESS
Cold Spring Harbor, New York

Molecular Cloning

A LABORATORY MANUAL

THIRD EDITION

Front cover (paperback): The gene encoding green fluorescent protein was cloned from *Aequorea victoria*, a jellyfish found in abundance in Puget Sound, Washington State. This picture of a 50-mm medusa was taken on color film by flash photography and shows light reflected from various morphological features of the animal. The small bright roundish blobs in the photograph are symbiotic amphipods living on or in the medusa. The bright ragged area in the center is the jellyfish's mouth.

Bioluminescence from *Aequorea* is emitted only from the margins of the medusae and cannot be seen in this image. Bioluminescence of *Aequorea*, as in most species of jellyfish, does not look like a soft overall glow, but occurs only at the rim of the bell and, given the right viewing conditions, would appear as a string of nearly microscopic fusiform green lights. The primary luminescence produced by *Aequorea* is actually bluish in color and is emitted by the protein aequorin. In a living jellyfish, light is emitted via the coupled green fluorescent protein, which causes the luminescence to appear green to the observer.

The figure and legend were kindly provided by Claudia Mills of the University of Washington, Friday Harbor. For further information, please see Mills, C.E. 1999–2000. Bioluminescence of *Aequorea*, a hydromedusa. Electronic Internet document available at http://faculty.washington.edu/cemills/Aequorea.html. Published by the author, web page established June 1999, last updated 23 August 2000.

Back cover (paperback): A portion of a human cDNA array hybridized with a red fluor-tagged experimental sample and a green fluor-tagged reference sample. Please see Appendix 10 for details. (Image provided by Vivek Mittal and Michael Wigler, Cold Spring Harbor Laboratory.)

Library of Congress Cataloging-in-Publication Data

Sambrook, Joseph.
Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell.-- 3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 0-87969-576-5 (cloth) -- ISBN 0-87969-577-3 (pbk)
1. Molecular cloning--Laboratory manuals.
[DNLM: 1. Cloning, Molecular--Laboratory Manuals. QH 440.5 S187m 2001] I. Russell, David W. (David William), 1954- . II. Title.
QH442.2 .S26 2001
572.8--dc21

00-064380

10 9 8 7 6 5 4 3 2 1

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory Press, 10 Skyline Drive, Plainview, New York 11803-2500. Phone: 1-800-843-4388 in Continental U.S. and Canada. All other locations: (516) 349-1930. FAX: (516) 349-1946. E-mail: cshpress@cshl.org. For a complete catalog of all Cold Spring Harbor Laboratory Press publications, visit our World Wide Web Site http://www.cshl.org/

For orders from Europe, the Middle East, and Africa, British pound pricing is provided. Orders are fulfilled and shipped from Cold Spring Harbor Laboratory Press–Europe c/o Lavis Marketing, 73 Lime Walk, Headington, Oxford OX3 7AD, U.K. Phone: +44 (0) 1865 741541. FAX: +044 (0)1865 750079. E-mail: cshlpress.europe@cshl.org. World Wide Web Site: http://www.cshlpress.co.uk

Volume 2

Introduction to Protocols 5, 6, 7, 8, and 9

Northern Hybridization

NORTHERN HYBRIDIZATION IS USED TO MEASURE the amount and size of RNAs transcribed from eukaryotic genes and to estimate their abundance. No other method is capable of obtaining these pieces of information simultaneously from a large number of RNA preparations; northern analysis is therefore fundamental to studies of gene expression in eukaryotic cells.

Northern hybridization became part of the standard repertoire of molecular biology almost immediately after the first descriptions of the method were published (Alwine et al. 1977, 1979). Although many variations and improvements (e.g., please see Kroczek 1993) have been published during the succeeding 20 years, the basic steps in northern analysis remain unchanged:

- isolation of intact mRNA
- separation of RNA according to size through a denaturing agarose gel
- transfer of the RNA to a solid support in a way that preserves its topological distribution within the gel
- fixation of the RNA to the solid matrix
- hybridization of the immobilized RNA to probes complementary to the sequences of interest
- removal of probe molecules that are nonspecifically bound to the solid matrix
- detection, capture, and analysis of an image of the specifically bound probe molecules.

There are choices at every step during the process and new alternatives continually appear in the literature. It is impossible to distill from this ferment the "best" combination of methods that can be universally applied in all situations. However, the methods described in the next five protocols are extremely robust and have worked well in a wide variety of circumstances.

SEPARATION OF RNA ACCORDING TO SIZE

Electrophoresis through denaturing agarose gels is used to separate RNAs according to size and is the first stage in northern hybridization. In earlier times, methylmercuric hydroxide (Bailey and Davidson 1976) achieved some degree of popularity, particularly among the brave and foolhardy. Although unparalleled as a denaturing agent, methylmercuric hydroxide is both volatile and extremely toxic (Cummins and Nesbitt 1978) and is therefore no longer recommended. The following are the two methods most commonly used today to separate denatured RNAs for northern analysis.

- Electrophoresis of RNA denatured with glyoxal/formamide through agarose gels (Protocol 5) (Bantle et al. 1976; McMaster and Carmichael 1977; Goldberg 1980; Thomas 1980, 1983).
- Pretreatment of RNA with formaldehyde and dimethylsulfoxide, followed by electrophoresis through gels containing up to 2.2 M formaldehyde (Protocol 6) (Boedtker 1971; Lehrach et al. 1977; Rave et al. 1979).

The two systems have approximately the same resolving power (Miller 1987), and the technical problems with both of them have long since been solved. For example, recirculation of electrophoresis buffer is no longer required when separating glyoxylated RNA in agarose gels and staining of RNA with ethidium bromide is now possible. However, glyoxal and especially formaldehyde retain some disadvantages, including toxicity. The choice between the systems therefore depends on the relative weight of these disadvantages, which will vary from one laboratory to the next.

Many compounds other than glyoxal, formaldehyde, and methylmercuric hydroxide have been explored as denaturing agents for RNA during gel electrophoresis, but few of these have proven to be reliable in routine laboratory use. Guanidine thiocyanate is the only compound that may have advantages over formaldehyde or glyoxal (Goda and Minton 1995). When incorporated into an agarose gel at a final concentration of 10 mM, it maintains RNA in a denatured form. Electrophoresis may be carried out in standard TBE buffer and ethidium bromide may be incorporated in the gel. However, few laboratories have adopted the method, and at present, experience with this system is too limited for us to recommend that guanidine thiocyanate be used in place of glyoxal and fomaldehyde.

EQUALIZING AMOUNTS OF RNA IN NORTHERN GELS

Equalizing the amounts of RNA loaded into lanes of northern gels is a thorny problem when a number of different samples are to be compared. Several different approaches are possible and none of them perfect:

- ***Loading of equal amounts of RNA*** (usually 0.5–0.7 OD_{260} units) into each lane of the gel. rRNAs are the dominant components in preparations of total cellular RNA and contribute >75% of the UV-adsorbing material. Northern analysis of equal quantities of total RNA shows how the steady-state concentration of target mRNAs changes with respect to rRNA content of the cell (Alwine et al. 1977; de Leeuw et al. 1989). Unlike the transcripts of housekeeping genes (see below), there is no evidence that the levels of 18S or 28S rRNA vary significantly from one mammalian tissue or cell line to the next (e.g., please see Bhatia et al. 1994). In addition, rRNA can easily be detected in agarose gels by staining with ethidium bromide instead of a second round of hybridization with a specific probe.
- ***Normalizing samples according to their content of mRNAs*** of an endogenous, constitutively expressed housekeeping gene such as cyclophilin, β-actin, or glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Kelly et al. 1983). All three genes are expressed at moderately abundant levels (~0.1% of poly(A)$^+$ RNA or 0.003% of total cellular RNA). Variations observed in the intensity of the hybridization signal of the gene of interest are often expressed relative to one of these three housekeeping genes. However, it turns out that the levels of expression of housekeeping genes are not constant from one mammalian tissue to another nor from one cell line to another (e.g., please see Spanakis 1993; Bhatia et al. 1994). Alterations in the relative intensity of the hybridization signals between the housekeeping gene and the gene of interest may therefore result from changes in the level of transcription of either gene or both.

- ***Loading of equal amounts of poly(A)$^+$ RNA.*** The poly(A)$^+$ content of preparations of RNA can be compared by slot- or dot-blot hybridization to a radiolabeled poly(dT) probe (Harley 1987, 1988). Equivalent amounts of poly(A)$^+$ RNA can then be loaded into each lane of a northern gel. This is an attractive option because it measures changes in concentration of a specific mRNA relative to the total amount of gene transcripts in the cell.
- ***Using a synthetic pseudomessage as a standard.*** Several groups (e.g., please see Toscani et al. 1987; DuBois et al. 1993) have used RNAs synthesized in vitro as externally added standards to calibrate the expression of the gene of interest in different preparations of cellular RNA. The synthetic pseudomessage, which is engineered to be different in size from the natural message, is added in known amounts to samples at the time of cell lysis. The relative intensity of the hybridization signals obtained from the authentic and pseudomessages is used to estimate the expression of the endogenous gene of interest.

MARKERS USED IN GELS TO FRACTIONATE RNA

The size of an RNA of interest can be measured accurately only when markers of known molecular weight are included in the gel. Four types of markers are commonly used:

- ***RNA standards purchased from a commercial source.*** These standards are usually generated by in vitro transcription of cloned DNA templates of known length. As a consequence, the RNA standards are sometimes contaminated by template DNA and its associated plasmid sequences. Vector sequences present in the probe used in northern hybridization may hybridize to these remnants, generating on the autoradiogram either discrete bands or, more commonly, a smear where none should be.
- ***DNA standards purchased from a commercial source.*** Glyoxylated denatured DNAs and RNAs of equal length migrate at equal speeds through agarose gels. Small DNAs of known size can therefore be used as markers in this system. Once again, however, there is a chance that vector sequences present in the probe may hybridize with the standards. At times, this can be an advantage because the signals generated by the marker bands on the autoradiogram can be used directly to measure the size of the RNA of interest. DNA standards should not be used as markers on gels containing formaldehyde since RNA migrates through these gels at a faster rate than DNA of equivalent size (Wicks 1986).
- ***Highly abundant rRNAs (28S and 18S) within the RNA preparations under test.*** The sizes of these RNAs vary slightly from one mammalian species to another. 18S rRNAs range in size from 1.8 kb to 2.0 kb, whereas 28S RNAs range between 4.6 kb and 5.3 kb in length.
- ***Tracking dyes.*** In most denaturing agarose gel systems, bromophenol blue migrates slightly faster than the 5S rRNA, whereas xylene cyanol migrates slightly slower than the 18S rRNA.

MEMBRANES USED FOR NORTHERN HYBRIDIZATION

Transfer of electrophoretically separated DNA and RNA from gels to two-dimensional solid supports is a key step in northern hybridization. Initially, hybridization was carried out exclusively with RNA immobilized on activated cellulose papers (Alwine et al. 1977; Seed 1982a,b). However, it was soon realized that RNA denatured by glyoxal, formaldehyde, or methylmercuric hydroxide, like denatured DNA, binds tightly to nitrocellulose (Thomas 1980, 1983). For several years thereafter, nitrocellulose was the support of choice for northern hybridization.

Unfortunately, nitrocellulose is not an ideal matrix for solid-phase hybridization because its capacity to bind nucleic acids is low (~50–100 μg/cm^2) and varies according to the size of the RNA. In addition, the RNA becomes attached to nitrocellulose by hydrophobic rather than covalent interactions and therefore leaches slowly from the solid support during hybridization and washing at high temperatures. Finally, nitrocellulose membranes become brittle during baking under vacuum at 80°C, which is an integral part of the process to immobilize nucleic acids. The friable membranes cannot subsequently survive more than two to three cycles of hybridization and washing at high temperatures.

These problems have been solved by the introduction of various types of nylon membranes that bind nucleic acids irreversibly, are far more durable than nitrocellulose filters (Reed and Mann 1985), and can be repaired if damaged (Pitas 1989). Immobilized nucleic acids can therefore be hybridized sequentially to several different probes. Furthermore, because nucleic acids can be immobilized on nylon in buffers of low ionic strength, transfer of nucleic acids from gels to nylon can be carried out electrophoretically. This advantage can be useful when capillary or vacuum transfer is inefficient, for example, when small molecules of RNA are transferred from polyacrylamide gels.

Two types of nylon membranes are available commercially: unmodified (or neutral) nylon and charge-modified nylon, which carries amine groups and is therefore also known as positively charged or (+) nylon. Both types of nylon bind single- and double-stranded nucleic acids and retention is quantitative in solvents as diverse as water, 0.25 N HCl, and 0.4 N NaOH. Charge-modified nylon has a greater capacity to bind nucleic acids (see Table 7-3), but it has a tendency to give increased levels of background hybridization, which results, at least in part, from nonspecific binding of negatively charged phosphate groups in RNA to the positively charged groups on the surface of the polymer. However, this problem can usually be controlled by using increased quantities of blocking agents in the prehybridization and hybridization steps.

Nylon is a generic name for any long-chain synthetic polymer having recurring polyamide (–CONH–) groups. Nylons of different types are formed from various combinations of diacids, diamines, and amino acids. In the standard nomenclature, a single numeral (e.g., nylon 6) indicates the number of carbon atoms in a monomer. Two numbers (e.g., nylon 6,6 or 66) indicate a polymer formed from diamines and dibasic acids. The first number indicates the number of carbon atoms separating the nitrogen atoms of the diamine, and the second number indicates the number of straight chain carbon atoms in the dibasic amino acid.

Fiber 66, the original name of nylon, was developed in the 1930s by Wallace Carothers, a chemist working for DuPont (see Fenichell 1999). His discovery, which grew from a decade of research on the structure and assembly of long-chain polyamide polymers, should have been the capstone of his career, but instead was the catalyst to tragedy. Carothers, more a scientific aesthete than a twentieth century company man, became deeply depressed by the idea that he had discovered a material whose chief use seemed to be as a replacement for silk stockings. In 1937, a few days after filing his patent for Fiber 66, Carothers, just 41 years old, killed himself in a hotel room by swallowing cyanide. DuPont pressed ahead with the commercial development of Fiber 66 and, in a ceremony that would have been anathema to Carothers, dedicated the name nylon to the public domain at a Herald Tribune Forum in October of the following year. Stockings, of course, turned out to be just the first of a line of nylon products, some of which would surely have given Carothers great pleasure, including perhaps, nylon membranes for immobilizing nucleic acids.

Different brands of nylon membranes are available that vary in the extent and type of charged groups and the density of the nylon mesh. Comparisons of the efficiency of these membranes for northern blotting and hybridization under various conditions are published from time to time (e.g., please see Khandjian 1987; Rosen et al. 1990; Twomey and Krawetz 1990; Beckers et al. 1994). In addition, each manufacturer provides specific recommendations for the transfer of nucleic acids to their particular product. The instructions given in Protocols 6 through 8 (northern hybridization) and in Chapter 6, Protocols 8–10 (Southern transfer) work well in almost all circumstances, and in some cases, yield results that exceed the manufacturer's standard.

TABLE 7-3 Properties of Nylon Membranes Used for Immobilization of DNA and RNA

PROPERTY	NEUTRAL NYLON	CHARGED NYLON
Capacity (μg nucleic acid/cm^2)	~200–300	400–500
Size of nucleic acid required for maximal binding	>50 bp	>50 bp
Transfer buffer	low ionic strength over a wide range of pH	
Immobilization	baking for 1 hour at 70°C; no vacuum required *or* mild alkali *or* UV irradiation at 254 nm; damp membranes are generally exposed to 1.6 kJ/m^2; dried membranes require 160 kJ/m^2	
Commercial products	Hybond-N Gene-Screen	Hybond-N+ Zeta-Probe Nytran + Gene-Screen Plus

TRANSFERRING RNA FROM GELS TO SOLID SUPPORTS

The crucial step in northern analysis is the transfer of denatured RNA from the interstices of an agarose gel to the surface of a membrane. Transfer must be done in a way that not only preserves the distribution of the molecules along the length of the gel, but works efficiently for nucleic acids of quite different sizes. Over the years, many methods have been found to achieve these goals, including electroblotting, vacuum blotting, semidry blotting, and upward capillary blotting. In addition, several attempts have been made to avoid transfer completely by performing hybridization directly in the gel (e.g., please see Purrello and Balazs 1983; Tsao et al. 1983). However, it is not clear whether these techniques, which may require expensive pieces of equipment, are superior to the original method of upward capillary transfer (Southern 1975). Certainly, there does not seem to be any good reason to rush out and buy a vacuum blotting or electroblotting apparatus in the belief that it will significantly improve northern and Southern blots.

- ***Upward capillary transfer.*** The original simple and economical technique devised by Southern (1975) involves an overnight transfer of nucleic acids from gel to membrane in an upward flow of buffer (please see Figure 7-2). A major drawback is selective retention of large molecules of nucleic acid within the gel, which is caused by flattening, compression, and dehydration of the gel. This problem can be relieved (1) by using the thinnest gels possible, (2) by ensuring that the filter papers in immediate contact with the gel are thoroughly saturated with buffer before transfer begins, and (3) by partial hydrolysis of RNA by alkali (Reed and Mann 1985) before transfer. It is important that partial hydrolysis be used with moderation since overenthusiasm can generate fragments too short to bind efficiently to the membrane.

  Since 1975, the common practice has been to carry out upward capillary transfer for 16 hours or so. However, ascending transfer is now known to be almost complete after 4 hours (Lichtenstein et al. 1990), and we now recommend much shorter transfer times. A more serious problem with ascending transfer is the potential for some of the RNA to move from the gel in a descending direction counter to the flow of the buffer. This apparent anomaly occurs when the filter paper under the gel is not fully saturated with buffer. Fluid is then drawn from

the gel, carrying with it some of the nucleic acid. The problem can be ameliorated by ensuring that the bottom filter paper, like the top, is fully saturated with buffer and by working quickly to set up the remainder of the transfer system once the gel has been laid on the bottom filter.

- ***Downward capillary transfer.*** Descending transfer (please see Figure 7-3) does not cause flattening of the agarose gel and results in a faster transfer of nucleic acid. RNA molecules up to 8 kb in size, for example, are transferred with high efficiency within 1 hour at either neutral or alkaline pH (Chomczynski 1992; Chomczynski and Mackey 1994). The speed of downward capillary transfer therefore has particular advantage when carrying out alkaline blotting of RNA. Blotting of RNA for more than 4 hours significantly decreases the strength of the hybridization signal, presumably due to excessive hydrolysis of the RNA.

FURTHER INFORMATION ABOUT NORTHERN HYBRIDIZATION

Northern and Southern hybridizations have much in common, including, for example, the mechanics of hybridization, the types of probes, and the posthybridization processing of the membranes. All of these topics are discussed in depth in other areas within this manual. Signposts to this information are posted at relevant positions within the next five protocols.

Protein Science

Transmembrane helix predictions revisited

Chien Peter Chen, Andrew Kernytsky and Burkhard Rost

Protein Sci. 2002 11: 2774-2791
Access the most recent version at doi:10.1110/ps.0214502

Transmembrane helix predictions revisited

CHIEN PETER CHEN,[1] ANDREW KERNYTSKY,[1] AND BURKHARD ROST[1,2,3]

[1]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA
[2]Columbia University Center for Computational Biology and Bioinformatics (C2B2), New York, New York 10032, USA
[3]North East Structural Genomics Consortium (NESG), Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA

(RECEIVED May 5, 2002; FINAL REVISION July 22, 2002; ACCEPTED September 16, 2002)

Abstract

Methods that predict membrane helices have become increasingly useful in the context of analyzing entire proteomes, as well as in everyday sequence analysis. Here, we analyzed 27 advanced and simple methods in detail. To resolve contradictions in previous works and to reevaluate transmembrane helix prediction algorithms, we introduced an analysis that distinguished between performance on redundancy-reduced high- and low-resolution data sets, established thresholds for significant differences in performance, and implemented both per-segment and per-residue analysis of membrane helix predictions. Although some of the advanced methods performed better than others, we showed in a thorough bootstrapping experiment based on various measures of accuracy that no method performed consistently best. In contrast, most simple hydrophobicity scale-based methods were significantly less accurate than any advanced method as they overpredicted membrane helices and confused membrane helices with hydrophobic regions outside of membranes. In contrast, the advanced methods usually distinguished correctly between membrane-helical and other proteins. Nonetheless, few methods reliably distinguished between signal peptides and membrane helices. We could not verify a significant difference in performance between eukaryotic and prokaryotic proteins. Surprisingly, we found that proteins with more than five helices were predicted at a significantly lower accuracy than proteins with five or fewer. The important implication is that structurally unsolved multispanning membrane proteins, which are often important drug targets, will remain problematic for transmembrane helix prediction algorithms. Overall, by establishing a standardized methodology for transmembrane helix prediction evaluation, we have resolved differences among previous works and presented novel trends that may impact the analysis of entire proteomes.

Keywords: Sequence analysis; protein structure prediction; multiple alignments, predicting transmembrane helices; comparing genomes; bioinformatics; computational biology; proteomes

Supplemental material: See www.proteinscience.org.

Reprint requests to: Burkhard Rost, Department of Biochemistry and Molecular Biophysics, Columbia University, 650 W. 168 St., BB217, New York, NY 10032, USA; e-mail: rost@columbia.edu; fax: (212) 305-7932.

Abbreviations: A-Cid, normalized hydrophobicity scale for α-proteins (Cid 1992); Av-Cid, normalized average hydrophobicity scale (Cid 1992); Ben-Tal, hydrophobicity scale representing the free energy of transferring an amino acid from water into the center of the hydrocarbon region of a lipid bilayer (Kessel and Ben-Tal 2002); BIG, nonidentical merger of SWISS-PROT (Bairoch and Apweiler 2000) and TrEMBL (Bairoch and Apweiler 2000) and PDB (Berman et al. 2000); BLAST, fast sequence alignment method (Altschul and Gish 1996); Bull–Breese, Bull–Breese hydrophobicity scale (Bull 1974); DSSP, program assigning secondary structure (Kabsch and Sander 1983); Eisenberg, normalized consensus hydrophobicity scale (Eisenberg et al. 1984); EM, Solvation free energy (Eisenberg and McLachlan 1986); EVA, server automatically evaluating structure prediction methods (Eyrich et al. 2001a,b); Fauchere, hydrophobic parameter π from the partitioning of *N*-acetyl-amino-acid amides (Fauchere and Pliska 1983); GES, hydrophobicity property (Engelman et al. 1986; Prabhakaran 1990); Heijne, transfer free energy to lipophilic phase (von Heijne and Blomberg 1979); HMM, hidden Markov model; HMMTOP, hidden Markov model predicting transmembrane helices (Tusnady and Simon 1998); Hopp-Woods, Hopp-Woods hydrophilicity value (Hopp and Woods 1981); KD, Kyte–Doolittle hydropathy index (Kyte and Doolittle 1982); Lawson, transfer free energy (Lawson et al. 1984); Levitt, hydrophobic parameter (Levitt 1976); MaxHom, dynamic programming algorithm for conservation weight-based multiple sequence alignment (Sander and Schneider 1991); MEMSAT, dynamic-programming based prediction of transmembrane helices (Jones et al. 1994); META-PP, inter-

Helical membrane proteins challenge bioinformatics. Membrane proteins are crucial for survival. They constitute key components for cell–cell signaling, mediate the transport of ions and solutes across the membrane, and are crucial for recognition of self (Stack et al. 1995; Chapman et al. 1998; Le Borgne and Hoflack 1998; Chen and Schnell 1999; Hettema et al. 1999; Pahl 1999; Truscott and Pfanner 1999; Bauer et al. 2000; Ito 2000; Soltys and Gupta 2000; Thanassi and Hultlgren 2000). Furthermore, the pharmaceutical industry preferably targets membrane-bound receptors (Heusser and Jardieu 1997; Bettler et al. 1998; Moreau and Huber 1999; Saragovi and Gehring 2000; Sedlacek 2000). Despite their great biological and medical importance, we still have very little experimental information about their 3D structures: <1% of the proteins of known structure are membrane proteins. Fortunately, it is relatively easy to identify the location of membrane helices through low-resolution experiments. An expert-curated list of low-resolution experiments maintained by Steffen Möller and colleagues (Möller et al. 2000) considers information from C-terminal fusions with indicator proteins (McGovern et al. 1991; Hennessey and Broome-Smith 1993; Traxler et al. 1993; van Geest and Lolkema 2000) and from antibody-binding studies (Traxler et al. 1993; McGuigan 1994; Jermutus et al. 1998; Morris et al. 1998; Amstutz et al. 2001). Nevertheless, we only have low-resolution experimental information for <500 helical membrane proteins, and PDB (Berman et al. 2000) contains <50 sequence-unique protein chains with high-resolution helical membrane structures (Materials and Methods). These numbers contrast with the >7000 helical membrane proteins expected in humans alone (Wallin and von Heijne 1998; Krogh et al. 2001; Liu and Rost 2001). Thus, bioinformatics is challenged to help bridge the information gap between what we want and what we have.

Published estimates for membrane helix prediction questioned by recent analyses. Recently, a few groups have questioned the estimated levels of performance for membrane helix prediction methods. Möller, Croning, and Apweiler analyzed 14 prediction methods that did not use alignment information on a set of 188 proteins with experimentally known helices (Möller et al. 2000, 2001). They also applied the prediction methods to globular proteins and to signal peptides. The results indicated the following conclusions: (1) The best prediction method (TMHMM, transmembrane prediction using cyclic hidden Markov models) correctly predicts all membrane helices for 52%–69% of all proteins tested. (2) The best distinction between globular and membrane-helical proteins reaches levels of >97% for the globular proteins tested (TMHMM and SOSUI, hydrophobicity- and amphiphilicity-based transmembrane helix prediction). (3) On a set of 34 signal and transit peptide proteins, the best methods reached 98% (PHDhtm, profile-based neural network prediction of transmembrane helices) to 100% (ALOM2) accuracy in distinguishing these from membrane helices. (4) The best simple hydrophobicity index (KD, Kyte–Doolittle hydropathy index; Kyte and Doolittle 1982) correctly predicted all helices for 44% of all the proteins in a set for which HMMTOP (hidden Markov model predicting transmembrane helices; Tusnady and Simon 1998) reached only 43% accuracy. Another recent analysis was based on a set of 145 sequence-unique proteins (Ikeda et al. 2001). The researchers tested 10 prediction methods not using alignment information on their data set. In contrast to Möller et al., the investigators found that HMMTOP was not only much better than the KD hydrophobicity index, but that it was the most accurate prediction method, correctly predicting all membrane helices for ~68% of all proteins. Averaging over all 10 methods, the authors found the resulting consensus prediction ~10 percentage points more accurate than the best single method. The investigators also claimed that prediction accuracy is higher for prokaryotes than for eukaryotes. They speculated that they found different levels of accuracy than Möller et al. because they used different percentages of prokaryotic proteins in their data sets. Jayasinghe, Hristova, and White analyzed four prediction methods on two different sets of proteins with known membrane helix locations: (1) on 150 high-resolution structures from PDB, and (2) on 242 low-resolution proteins (Jayasinghe et al. 2001b). The researchers found that the results between the high- and low-resolution sets differed marginally and reported that the best

net service allowing access to a variety of bioinformatics tools through one single interface (Eyrich and Rost 2000); Nakashima, normalized composition of membrane proteins (Nakashima et al. 1990); PDB, Protein Data Bank of experimentally determined 3D structures of proteins (Bernstein et al. 1977; Berman et al. 2000); PHDhtm, profile-based neural network prediction of transmembrane helices (Rost 1996; Rost et al. 1996b); PHDpsihtm, divergent profile (PSI-BLAST)-based neural network prediction 2002); PSI-BLAST, position-specific iterated database search (Altschul et al. 1997); Radzicka, transfer free energy from 1-octanol to water (Radzicka and Wolfenden 1988); Roseman, solvation-corrected side-chain hydropathy (Roseman 1988); SignalP, signal peptide prediction (Nielsen et al. 1997a); SOSUI, hydrophobicity- and amphiphilicity-based transmembrane helix prediction (Hirokawa et al. 1998); SPLIT, transmembrane helix prediction (Juretic et al. 1998); Sweet, optimal matching hydrophobicity (Sweet and Eisenberg 1983); SWISS-PROT, database of protein sequences (Bairoch and Apweiler 2000); TM, transmembrane; TMAP, alignment-based prediction of transmembrane helices (Persson and Argos 1996); TMH, transmembrane helix; TMHMM, transmembrane prediction using cyclic hidden Markov models (Sonnhammer et al. 1998; Krogh et al. 2001); TMpred, prediction of transmembrane helices (Hofmann and Stoffel 1993); TopPred2, hydrophobicity-based membrane helix prediction (von Heijne 1992; Cserzö et al. 1997); TrEMBL, translation of the EMBL-nucleotide database coding DNA to protein sequences (Bairoch and Apweiler 2000); Wolfenden, hydration potential (Wolfenden et al. 1981); WW, Wimley–White hydrophobicity scale-based method (Wimley et al. 1996a,b; White and Wimley 1999; White 2001).

Terminology: Advanced prediction methods, all methods that do not exclusively use a hydrophobicity scale; simple prediction methods, membrane prediction methods exclusively based on hydrophobicity scales.

Formula abbreviations: htm, transmembrane helix; T, residue in transmembrane helix; N, nonmembrane residue.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0214502.

methods (PHDhtm and HMMTOP) correctly predict >93%–97% of all helices. This group has also proposed a method based on a novel entropy-based hydrophobicity scale, namely, the Wimley–White scale (WW, Wimley–White hydrophobicity-scale-based method), which is claimed to correctly predict 99% of all membrane helices (Jayasinghe et al. 2001a). One major problem of hydrophobicity-based methods appears to be the poor distinction between membrane and globular proteins (Edelman 1993; Jones et al. 1994; Rost et al. 1995, 1996b; Jayasinghe et al. 2001a; Möller et al. 2001).

Problems with previous analyses. Previous analyses were limited in various ways. (1) Performance on high- and low-resolution data sets was distinguished by neither the Möller nor the Ikeda groups, although it seemed that performance differed between the two (Jayasinghe et al. 2001b). (2) The redundancy in data sets resulting from many copies of very similar proteins was not reduced by the Möller or Jayasinghe groups. However, such bias is known to create problems when estimating prediction methods (Rost and Sander 1993; Rost et al. 1995, 1996b; Rost 2002). (3) Neither Möller et al. nor Ikeda et al. tested any method based on alignment information, although such methods are known to be more accurate (Rost and Sander 1993; Persson and Argos 1994; Neuwald et al. 1995; Rost et al. 1995; Rost 1996; Johnson and Church 1999). (4) No group explored per-residue—along with per-segment—based measures for prediction accuracy. Instead, all groups focused on one particular definition of prediction accuracy; no two groups applied the same definition. (5) No group established levels for significant differences between methods. This makes it impossible to conclude whether or not differences between any two methods are relevant. In general, levels of significant differences typically depend on the data sets and the scores used (Eyrich et al. 2001; Rost and Eyrich 2001; Marti-Renom et al. 2002). (6) Only Möller and coworkers tested proteins with signal peptides; however, their analysis was restricted to a small set of 34 proteins with known signal peptides. (7) No group analyzed more than 14 prediction methods. (8) Generally, prediction accuracy differs significantly between proteins used to develop a method and proteins never seen by a method (Moult et al. 1995, 1997, 1999). For membrane proteins, this effect is very difficult to estimate because few high-resolution structures of membrane proteins are added over a course of a year. Although Möller et al. tried to estimate this effect by analyzing only proteins not used for developing a method, they did not rule out that the proteins tested in the category "not known to the method" were similar to proteins used for development. Surprisingly, Möller et al. found most methods to perform better on proteins not used for development. Given how prediction methods are developed, it is very unlikely that this result holds in general. Either the differences are not significant, or the data sets were not representative (or both).

To resolve these limitations and to standardize membrane helix prediction performance comparisons, we have presented an analysis that distinguished between performance on redundancy-reduced high- and low-resolution data sets, established thresholds for significant differences in performance by introducing a bootstrap experiment, and implemented both per-segment and per-residue analysis of membrane helix predictions. Additionally, we analyzed more methods (8 publicly available advanced prediction methods and 19 different hydrophobicity scales). In particular, we included alignment-based prediction methods. Furthermore, we tested membrane helix prediction methods on a large, representative set of 1418 unique signal peptides and 616 unique globular protein folds taken from SCOP (Lo Conte et al. 2002). Although we confirmed many previous findings, overall our results differed greatly in detail from previous publications.

Results

Accuracy in predicting membrane helices

Prediction methods not significantly less accurate than low-resolution experiments! We compared the membrane annotations for 13 proteins for which we had both low-resolution and high-resolution data available. Whereas ~94%–96% of the helices agreed between the two experimental methods, for only 11 of the 13 proteins did all helices overlap between the two experimental methods (Table 1). Also, the two methods agreed on only 82% of all residue assignments (Table 1, Q_2, percentage of correctly predicted residues in two states: membrane helix and non-membrane helix). A detailed comparison of the percentage of identically assigned membrane-helical residues confirmed that for most cases, the differences arose from the longer segments observed in the high-resolution data ($Q_{2T}^{\%obs} < Q_{2T}^{\%prd}$, where $Q_{2T}^{\%obs}$ is the percentage of all observed TMH helix residues that are correctly predicted and $Q_{2T}^{\%prd}$ is the percentage of all predicted TMH helix residues that are correctly predicted). Assuming that the high-resolution data were correct, we can interpret the low-resolution data as an experimental prediction of transmembrane helices. Surprisingly, most prediction methods performed as well as the low-resolution experiments (Table 1). In fact, in terms of almost all measures for accuracy, we could find one method that numerically agreed more with the high-resolution data than the low-resolution experiment. However, given the small size of the data set, this statement ignored the error margins in the estimate for accuracy.

Simple hydrophobicity-based predictions were less accurate than advanced methods. Of the methods that only used hydrophobicity scales for prediction, none detected all membrane helices correctly for >70% of the high-resolution

Table 1. *Accuracy of low-resolution experiments and predictions*

	Per-segment accuracy[b]				Per-residue accuracy[c]				
Method[a]	Q_{ok}	$Q_{htm}^{\%obs}$	$Q_{htm}^{\%prd}$	TOPO	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%prd}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%prd}$
ERROR[d]	±16	±10	±10	±16	±6	±9	±9	±7	±7
LOW-RES	84	98	96	75	82	70	90	92	71
DAS	55	96	91		*69*[e]	*46*	91	94	58
HMMTOP2	93	99	99	62	78	67	88	85	66
PHDhtm08	83	98	98	64	79	74	77	85	82
PHDhtm07	85	98	98	64	79	74	77	85	82
PHDpsiHtm08	92	98	100	92	79	74	81	88	83
PRED-TMR	*44*	80	93		71	53	81	91	61
SOSUI	77	90	92		78	66	79	84	67
TMHMM1	77	89	92	53	80	66	82	87	68
TopPred2	78	96	99	61	76	65	87	83	65
WW	52	88	87		72	68	*66*	*64*	67

[a] Methods: see abbreviations at begin of article.
[b] Per-segment accuracy: Q_{ok} gives the percentage of proteins for which all TM helices are predicted correctly (eq. 4), $Q_{htm}^{\%obs}$ the percentage of all observed helices that are correctly predicted (eq. 2), $Q_{htm}^{\%prd}$ is the percentage of all predicted helices that are correctly predicted (eq. 3), TOPO the percentage of proteins for which the topology (orientation of helices) is correctly predicted (eq. 4, not: empty for methods that do not predict topology).
[c] Per-residue accuracy: Q_2 is the percentage of correctly predicted residues in two-states: membrane helix/nonmembrane helix (eq. 6), $Q_{2T}^{\%obs}$ the percentage of all observed TMH helix residues that are correctly predicted (eq. 7), $Q_{2T}^{\%prd}$ the percentage of all predicted TMH helix residues that are correctly predicted (eq. 8), $Q_{2N}^{\%obs}$ the percentage of all observed non-TMH helix residues that are correctly predicted, and $Q_{2N}^{\%prd}$ the percentage of all predicted non-TMH helix residues that are correctly predicted.
Note of caution: this data set of 13 proteins was too small to rank the prediction methods in any way!
Data set: 13 high-resolution membrane helical proteins from PDB for which we found low-resolution experimental information in old versions of SWISS-PROT (labeled by LOW-RES). Note that the topology assessment was based on only 8 of the 13 proteins for which we had this information.
[d] ERROR: The estimates for per-segment accuracy resulted from a bootstrap experiment with $M = 100$ and $K = 6$ (Fig. 5); the estimates for per-residue accuracy were obtained according to equation 11.
[e] Numbers in italics: 2 standard deviations difference from baseline LOW-RES.

proteins (Table 2, Q_{ok}, percentage of proteins for which all TM helices are predicted correctly). However, most methods correctly identified >90% of all observed membrane helices (Table 2, $Q_{htm}^{\%obs}$, percentage of all observed helices that are predicted correctly). In fact, measured by this score alone, most simple hydrophobicity-based methods appeared more accurate than many advanced prediction methods, but this success was achieved by overpredicting membrane helices (Table 2, $Q_{htm}^{\%prd} < Q_{htm}^{\%obs}$, where $Q_{htm}^{\%prd}$ is the percentage of all predicted helices that are predicted correctly). Encouragingly, >80% of the helices predicted by most methods were correct (Table 2, $Q_{htm}^{\%prd}$). Unfortunately, the real problem with the simple methods was that they did not correctly predict the nonmembrane regions as apparent in levels of <70% correctly predicted residues (Table 2, Q_2). Note that we implemented all simple hydrophobicity scales by using the algorithm proposed by the White group (Jayasinghe et al. 2001a). To ensure that this optimized or at least did not penalize membrane protein prediction for some hydrophobicity scales, we also tested the thresholds suggested in the original publications for the GES (hydrophobicity property; Engelman et al. 1986; Prabhakaran 1990) and KD scales (Kyte and Doolittle 1982). Interestingly, the originally proposed thresholds decreased prediction accuracy (Supplementary Table 1; available online at http://www.proteinscience.org).

Most advanced predictions were correct. All advanced prediction methods correctly identified all helices for most high-resolution proteins (Table 2, Q_{ok}). In contrast, the only two methods we found to also accurately predict the orientation of the helices, that is, the topology, most often were TopPred2 (hydrophobicity-based membrane helix prediction) and HMMTOP2 (Table 2, TOPO, percentage of proteins for which the topology is correctly predicted). Note that HMMTOP2 was developed using all the 36 high-resolution chains for which we compiled the results. On the other hand, TopPred2 used only four of the 36 chains when it was developed. All methods tested correctly predicted >70% of the residues in either of the two states, TMH (T) and non-TMH (N, Table 2, Q_2). However, all methods significantly underpredicted residues in membrane helices (Table 2, $Q_{2T}^{\%obs} < Q_{2T}^{\%prd}$).

No single advanced method best by all scores. The set of 36 high-resolution proteins was small enough to require extreme caution in ranking methods based on numerical differences. When comparing pairwise ranks of the methods

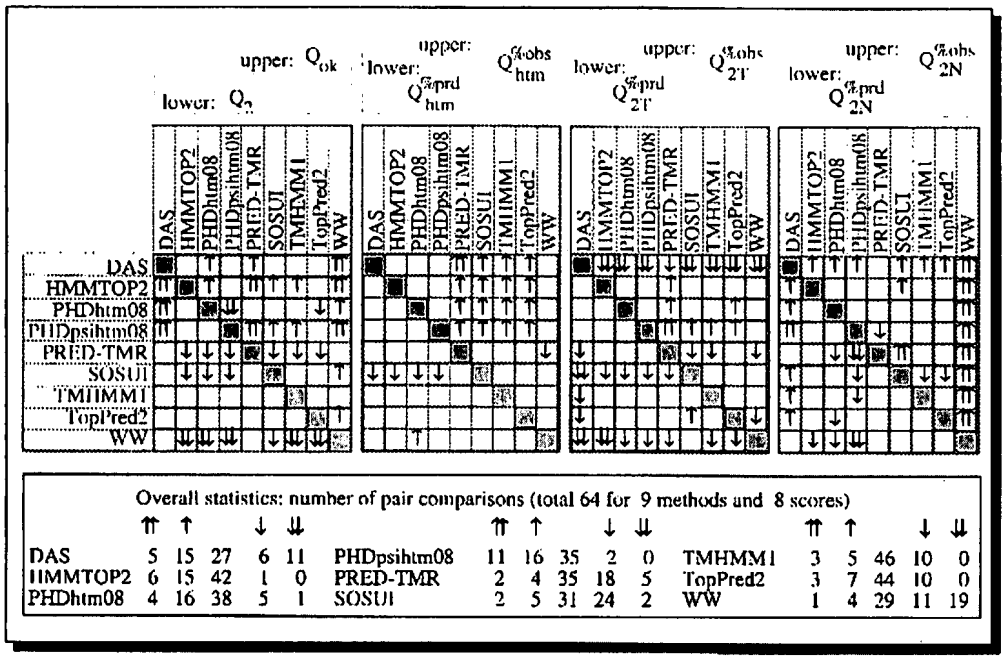
Table 2. *Accuracy of prediction methods for high-resolution set*

	Per-segment accuracy				Per-residue accuracy				
Method	Q_{ok}	$Q_{htm}^{\%obs}$	$Q_{htm}^{\%prd}$	TOPO	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%prd}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%prd}$
ERROR	±10	±8	±10	±9	±3	±7	±8	±6	±6
DAS	79	99	96		*72*	*48*	94	96	*62*
HMMTOP2	83	99	99	61	80	*69*	89	88	71
PHDhtm08	64	*77*	*76*	54	78	76	82	84	79
PHDhtm07	69	83	81	50	78	76	82	84	79
PHDpsihtm08	84	99	98	66	80	76	83	86	80
PRED-TMR	*61*	84	90		76	*58*	85	94	*66*
SOSUI	71	88	86		75	*66*	*74*	*80*	69
TMHMM1	71	90	90	*45*	80	*68*	81	89	72
TopPred2	75	90	90	54	77	*64*	83	90	69
KD	*65*	94	89		*67*	79	*66*	*52*	*67*
GES	*64*	97	90		*71*	74	*72*	*66*	69
Ben-Tal	*60*	*79*	89		*72*	*53*	80	95	*63*
Eisenberg	*58*	95	89		*69*	77	*68*	*57*	68
Hopp-Woods	*56*	93	86		*62*	80	*61*	*43*	*67*
WW	*54*	95	91		*71*	71	*72*	*67*	*67*
Av-Cid	*52*	93	83		*60*	83	*58*	*39*	72
Roseman	*52*	94	83		*58*	83	*58*	*34*	*66*
Levitt	*48*	91	84		*59*	80	*58*	*38*	*67*
A-Cid	*47*	95	83		*58*	80	*56*	*37*	*66*
Heijne	*45*	93	82		*61*	85	*58*	*34*	*64*
Bull-Breese	*45*	92	82		*55*	85	*55*	*27*	*66*
Sweet	*43*	90	83		*63*	83	*60*	*43*	69
Radzicka	*40*	93	79		*56*	85	*55*	*26*	*63*
Nakashima	*39*	88	83		*60*	84	*58*	*36*	*63*
Fauchere	*36*	92	80		*56*	84	*56*	*31*	*65*
Lawson	*33*	86	79		*55*	84	*54*	*27*	*63*
EM	*31*	92	*77*		*57*	85	*55*	*28*	*64*
Wolfenden	*28*	*43*	*62*		*62*	*28*	*56*	97	*56*

Data set: 36 high-resolution membrane helical proteins from PDB; Note: We had reliable information about topology for only 35 of the 36 proteins.
Abbreviations as in Table 1.
Methods, hydrophobicity scales: see the abbreviations footnote at the beginning of the article for the advanced methods, and the list of hydrophobicity scales in the Materials and Methods section for the hydrophobicity scales. The advanced methods are sorted by alphabet, the simple hydrophobicity-based methods according to the Q_{ok} score.
ERROR: the estimates for per-segment accuracy resulted from a bootstrap experiment with $M = 100$ and $K = 18$ (Fig. 5); the estimates for per-residue accuracy were obtained according to equation 11.
Numbers in italics: two standard deviations below the numerically highest value in each column.
Note of caution: all methods are tested on the same set of proteins. However, the numbers are not from a cross-validation experiment, that is, some methods may have used some of the proteins for training. Generally, newer methods are more likely to be overestimated than older ones.

according to various scores, we found that no advanced method performed consistently best, and none consistently worst (Fig. 1). Interestingly, TMHMM1 and TopPred2 appeared to be the most representative methods in that the scores for these methods were most often indistinguishable from all other advanced methods in pairwise comparisons. In contrast, DAS appeared to be most unique in that it was often better and often worse than all other methods. Three methods were clearly more often worse than better: WW (5 times better/30 times worse), PRED-TMR (6/23), and SOSUI (7/26). Three methods were clearly more often better than worse: HMMTOP2 (21 times better/1 time worse), PHDpsihtm08 (divergent profile-based neural network prediction of transmembrane helices) (27/2), and PHDhtm08 (20/6).

Performance on low-resolution data set: distinct differences. The low-resolution set was considerably larger (165 proteins) than the high-resolution set (36 chains). Nevertheless, we could still not find any method that performed consistently better than all the others (Table 3). Most methods reached better per-segment scores for the high- than for the low-resolution data. The opposite was the case for per-residue scores as they were consistently higher for the low-resolution proteins. Most surprising may be the significant differences between the two data sets in terms of the percentage of proteins for which all helices were correctly pre-



Overall statistics: number of pair comparisons (total 64 for 9 methods and 8 scores)

	⇈	↑		↓	⇊		⇈	↑		↓	⇊		⇈	↑		↓	⇊
DAS	5	15	27	6	11	PHDpsihtm08	11	16	35	2	0	TMHMM1	3	5	46	10	0
HMMTOP2	6	15	42	1	0	PRED-TMR	2	4	35	18	5	TopPred2	3	7	44	10	0
PHDhtm08	4	16	38	5	1	SOSUI	2	5	31	24	2	WW	1	4	29	11	19

Fig. 1. Pairwise comparison of methods. For all high-resolution results compiled in Table 2, we show the pairwise comparison for eight different scores and nine methods. Differences by more than one (two) standard error(s) are marked by one (two) arrow(s). Empty boxes indicate that the difference between the respective scores of the two methods is not significant. For example, DAS is two standard errors better than WW in terms of the number of correctly predicted proteins (Q_{ok}), whereas HMMTOP2 is two standard errors better than DAS in terms of the overall per-residue accuracy (Q_2). The lower table summarizes the respective counts of pair-comparisons for which a particular method is better or worse than the others. TopPred2 and TMHMM1 appear to be the most neutral method (44 and 46 times indistinguishable), whereas DAS seems the most unique method in that it is often better than the others and equally often worse. Note: only DAS, PHDhtm08, PHDpsihtm07, and TopPred2 did not use most of the proteins tested to optimize prediction accuracy; thus, the results for all the other methods are likely to be overestimates.

dicted for the old methods DAS and TopPred2 (Q_{ok} in Tables 2 and 3). Even more stunning was the extremely poor performance of most simple methods using only hydrophobicity scales for the prediction. Interestingly, for the hydrophobicity scales, the two newest ones (WW and Ben-Tal; hydrophobicity scale representing the free energy of transferring an amino acid from water into the center of the hydrocarbon region of a lipid bilayer) performed best overall on the data from low-resolution experiments.

Most errors were under- or overpredictions of one TMH. The good news was that all methods predicted the number of membrane helices correctly for most proteins (Fig. 2). However, this number differed significantly between the high- (71%) and the low-resolution data (56%). The majority of deviations were to predict one helix too few or one too many (68% for high; 64% for low-resolution, Fig. 2, center). Interestingly, the errors were rather symmetric for the low-resolution set, whereas they were substantially asymmetric for the high-resolution data. We could not find any significant correlation between the number of membrane helices and the errors of a particular method (data not shown). However, this may be largely owing to the few high-resolution structures in our data set.

Accuracy lower for proteins with more than five TMH's. For proteins with five or fewer membrane helices, the average over all advanced methods exceeded 80% (Q_{ok}, eq. 4) for the high-resolution data and 60% for the low-resolution data (Fig. 3). However, prediction accuracy dropped significantly for proteins with more than five helices to values from 33%–36% (Fig. 3). Why are proteins with less than five TMH's so different from proteins with more than six TMH's? Answers to this question remain speculative.

Most proteins and most helices correctly predicted by one of the methods. None of the high-resolution helices has been consistently mispredicted by all programs. However, this may reflect that the more recent methods used all these proteins for training. In contrast, three transmembrane helices from three proteins of the low-resolution set were not identified by any of the methods: (1) The C4-dicarboxylate transport protein from *Rhizobium meliloti* (SWISS-PROT ID dcta_rhime; helix from residues 282–300, sequence ALPGLMNKMEKAGCKRSVV) has a relatively hydrophobic sequence, but it has a polar stretch of residues, NKMEK, in the middle of the helix. The gene fusion constructs were not always created with the reporter gene present in the predicted loop regions (Jording and Puhler 1993).

Table 3. *Accuracy of prediction methods for low-resolution set*

	Per-segment accuracy				Per-residue accuracy				
Method	Q_{ok}	$Q_{htm}^{\%obs}$	$Q_{htm}^{\%prd}$	TOPO	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%prd}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%prd}$
ERROR	±9	±5	±5	±9	±2	±4	±4	±2	±2
DAS	*39*	93	*81*		86	*65*	85	97	*84*
HMMTOP2	66	94	93	79	90	85	83	*91*	*91*
PHDhtm08	57	86	86	68	87	83	*75*	*90*	94
PHDhtm07	56	85	86	72	87	83	*75*	*90*	94
PHDpsiHtm08	67	95	94	67	89	87	*77*	*92*	96
PRED-TMR	58	92	93		90	*78*	86	94	*89*
SOSUI	*49*	88	86		88	*79*	*72*	*88*	*90*
TMHMM1	72	91	92	85	90	83	80	*91*	92
TopPred2	*48*	84	*79*	*59*	88	*74*	*71*	*93*	*89*
Ben-Tal	*35*	*79*	90		87	*67*	83	95	*85*
Wolfenden	*29*	*56*	*82*		*80*	*47*	*76*	97	*79*
WW	*27*	90	75		*81*	83	*59*	*77*	89
GES	*23*	93	*68*		*78*	87	*53*	*72*	*91*
Eisenberg	*20*	90	*63*		*72*	89	*47*	*63*	*91*
KD	*13*	88	*59*		*63*	91	*42*	*50*	*91*
Heijne	*11*	89	*55*		*51*	91	*35*	*33*	89
Hopp-Woods	*11*	87	*58*		*54*	90	*36*	*38*	*88*
Sweet	*11*	87	*59*		*58*	88	*38*	*44*	*87*
Av-Cid	*10*	87	*58*		*53*	89	*36*	*38*	*87*
Roseman	*9*	89	*56*		*48*	91	*34*	*30*	*88*
Levitt	*9*	88	*56*		*49*	91	*35*	*32*	*88*
Nakashima	*9*	88	*56*		*50*	90	*35*	*34*	*87*
A-Cid	*8*	87	*57*		*52*	89	*35*	*36*	*87*
Lawson	*8*	86	*57*		*43*	89	*32*	*24*	*83*
Radzicka	*6*	87	*56*		*41*	91	*32*	*21*	*85*
Bull-Breese	*6*	86	*56*		*40*	91	*32*	*20*	*83*
EM	*5*	89	*56*		*41*	91	*32*	*21*	*85*
Fauchere	*5*	87	*56*		*43*	91	*33*	*23*	*86*

Data set: 165 low-resolution membrane helical proteins from SWISS-PROT (Möller et al. 2000).
Note: We had reliable information about topology only for 140 of the 165 proteins.
Abbreviations as in Table 2. The advanced methods are sorted by alphabet, the simple hydrophobicity-based methods according to the Q_{ok} score.
Numbers in italics: two standard deviations below the numerically highest value in each column.
Note of caution: all methods are tested on the same set of proteins. However, the numbers are not from a cross-validation experiment, that is, some methods may have used some of the proteins for training. Generally, newer methods are more likely to be overestimated than older ones. In particular, DAS, the PHD methods, and TopPred2 used only a small subset of these proteins for setting up the method, whereas HMMTOP2 used most.

In some cases, the reporter gene was present in the predicted membrane regions. This is a problem because it may alter the topological placement of the reporter gene with respect to the membrane. In addition, gene fusion constructs were not made for each loop region because reporter genes were introduced at random. Hence, each loop was not tested, which included loops for helix 282–300, for its topological placement. Hence, the experimental evidence for this membrane helix (282–300) was weak, at best. (2) The Haemolysin Secretion ATP-Binding Protein (HlyB) from *Escherichia coli* (hlyb_ecoli, residues 38–51, sequence GTGL GLTSWLLAAK) is an integral membrane protein. However, the particular membrane helix missed appears very short. The other seven membrane helices of HlyB are at least 20 residues long. However, some authors have claimed that membrane-spanning helices may be as short as 10 residues long (Lewis et al. 1990). The experimental evidence for hylb_ecoli had similar problems as that for dcta_rhime: The experimentalists found it difficult to identify membrane-spanning regions through predictions (Wang et al. 1991). This was caused by the high proportion of hydrophilic residues in the N-terminal portion of HlyB. Consequently, the authors did not know where to insert their reporter gene, which in this case was β-lactamase. Thus, they randomly inserted the reporter gene. Additionally, topological models identify the short stretch as loop (Wang et al. 1991; Gentschev and Goebel 1992). (3) Like all other problematic cases, the Mitochondrial brown fat uncoupling protein 1 from *Rattus norvegicus* (ucp1_rat, residues 178–194, sequence PNLMRNVIINCTELVTY) has transmembrane

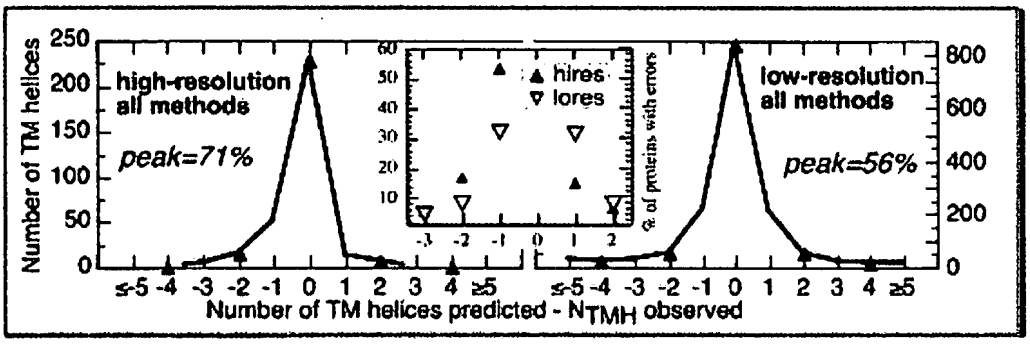


Fig. 2. Over- and underprediction of membrane helices. All methods (*top* panel): For all methods and all proteins in the high- and low-resolution sets, the difference between the number of membrane helices predicted and observed is shown. Although the two distributions appear rather similar, the higher symmetry in the low-resolution graph hid that the percentages with no difference were quite different: 71% for the high-resolution data and 56% for the low-resolution data. The inset (*center*) underlined the observation that the majority of errors were due to under- or overpredicting one helix.

regions that contain many polar residues. For this protein, the experimentalists stated that their data did not suffice to strongly conclude that residues 178–194 are in a membrane helix (Miroux et al. 1993).

No significant difference in performance for prokaryotic and eukaryotic proteins. We compared the performance of each method for eukaryotic and prokaryotic proteins. Most methods did not consistently perform better for both the high- and low-resolution data (Table 4, ΔQ_{ok}). In fact, the trends differed greatly between both data sets, and for different measures of prediction accuracy. Whereas prokaryotic proteins were predicted more accurately in terms of per-segment measures for the high-resolution data sets, the opposite was the case for most methods when compared on the low-resolution set. Only four methods had a similar trend in Q_{ok}: PRED-TMR predicted eukaryotic proteins more accurately; SOSUI, TopPred2, and WW predicted prokaryotic proteins more accurately for both sets. However, none of the values exceeded two times the estimated error, that is, none was statistically very significant. All methods predicted topology (ΔTOPO) better for the prokaryotic proteins in the high-resolution set and for the eukaryotic proteins in the low-resolution set. When measuring prediction accuracy in terms of per-residue performance (ΔQ_2), we could not find any significant difference between prokaryotic and eukaryotic proteins; all methods did slightly

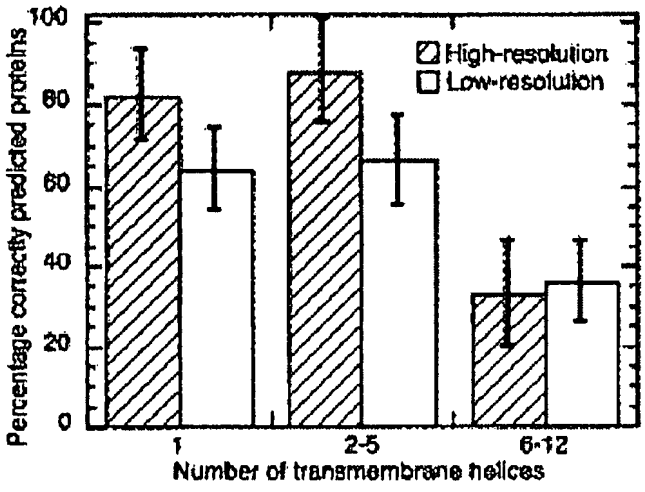


Fig. 3. Proteins with many helices predicted less accurately. We binned the results for all advanced methods according to the number of observed membrane helices such that the three classes contained similar numbers of proteins (*X*-axis). Accuracy (*Y*-axis) is measured in terms of the percentage of proteins for which all helices are correctly predicted (Q_{ok}). Both, for the high- and the low-resolution data, proteins with more than five membrane helices were predicted at significantly lower levels of accuracy.

Table 4. *Difference between eukaryotic and prokaryotic membrane proteins*

	Difference in accuracy eukaryotes vs. prokaryotes					
	High-resolution			Low-resolution		
Method	ΔQ_{ok}	ΔTOPO	ΔQ_2	ΔQ_{ok}	ΔTOPO	ΔQ_2
ERROR	±14	±12	±20	±18	±6	±18
DAS	4		4	−16		8
HMMTOP2	−9	*−31*	2	13	6	9
PHDhtm08	−24	−14	3	10	*39*	10
PHDhtm07	−11	−6	3	10	*39*	10
PHDpsiHtm08	−20	*−32*	3	13	*32*	8
PRED-TMR	5		5	11		7
SOSUI	−8		1	−18		5
TMHMM1	−20	*−39*	0	6	*12*	5
TopPred2	−12	−18	0	−12	*−32*	7
WW	−6		4	−12		5

Data set: eukaryotic proteins: 19 in high-resolution set, 73 in low-resolution set; prokaryotic proteins: 17 in high-resolution set, 87 in low-resolution set.
Accuracy: levels of accuracy given are the differences in the averages over all eukaryotic proteins minus the averages over all prokaryotic proteins.
Number in italics: values that are ± two standard deviations from a difference of 0.

better for eukaryotic proteins for both high- and low-resolution data. Nevertheless, because of the lack of consistent direction of the difference and the lack of statistical significance, our data did not support the previously published conclusion that either prokaryotic or eukaryotic proteins were predicted more accurately.

Accuracy of distinguishing between membrane and other proteins

Few false positives: best methods found few membrane helices in globular proteins. Most advanced methods correctly distinguished between membrane and globular proteins (Table 5). The best methods confused between the two types of proteins for <4% of all globular proteins tested (Table 5). DAS had the highest error rate of the advanced methods (16% false positives), which was surprising given that DAS tended to underpredict residues in membrane helices. In contrast to the advanced methods, the simple methods distinguished only poorly between membrane and globular proteins. The two exceptions were the old scale from Wolfenden (hydration potential; Wolfenden et al. 1981) and the new one from Ben-Tal (Kessel and Ben-Tal 2002). The latter also predicted membrane proteins rather accurately (Tables 2 and 3). However, most methods found helices in >90% of all the globular proteins.

Table 5. *Confusion of membrane and globular proteins*

		False negatives (%)	
Method	False positives (%)	High-resolution	Low-resolution
ERROR	±2	±9	±3
SOSUI	1	8	4
TMHMM1	1	8	4
Wolfenden	2	*39*	*13*
PHDpsihtm	2	3	*8*
PHDhtm08	2	*19*	*23*
Ben-Tal	3	11	4
PHDhtm07	3	14	*16*
PRED-TMR	4	8	1
HMMTOP2	*6*	0	1
TopPred2	*10*	8	*11*
DAS	*16*	0	0
WW	*32*	0	0
GES	*53*	0	0
Eisenberg	*66*	0	0
KD	*81*	0	0
Sweet	*84*	0	0
Hopp-Woods	*89*	0	0
Nakashima	*90*	0	0
Heijne	*92*	0	0
Levitt	*93*	0	0
Roseman	*95*	0	0
A-Cid	*95*	0	0
Av-Cid	*95*	0	0
Lawson	*98*	0	0
FM	*99*	0	0
Fauchere	*99*	0	0
Bull-Breese	*100*	0	0
Radzicka	*100*	0	0

Data set: 616 high-resolution globular proteins from PDB (for false positives, i.e., the test whether or not the methods incorrectly predict membrane helices in globular proteins). The membrane sets are identical to those given in Table 2 (high-resolution) and Table 3 (low-resolution).
Methods are sorted by the accuracy in correctly rejecting globular protein (false positives).
Numbers in italics: two standard errors below the lowest confusion rate.

Few false negatives. Most methods find all membrane proteins. Although most hydrophobicity scales detected membrane helices in >90% of the globular proteins, they detected all membrane proteins as such. The exceptions were the two scales that were best in rejecting globular proteins: Wolfenden and Ben-Tal (Table 5). Similarly, PHDhtm08 misclassified only 2% of the globular proteins, but also missed ~20% of the membrane proteins. The only methods that misclassified <10% of the globular proteins and overlooked <10% of the membrane proteins were: SOSUI, TMHMM1, PHDpsihtm, PRED-TMR, and HMMTOP2 (Table 5).

Signal peptides falsely predicted to be membrane helices by most methods. Even the advanced methods had high error rates for signal peptides (Table 6). In fact, one of the most accurate rejections of signal peptides was achieved by the simple method solely using the Wolfenden (Wolfenden et al. 1979) hydrophobicity scale (26% errors). Many of the false predictions were at the very beginning of the respective secreted proteins. Thus, we tested the following simple expert rule: delete all membrane helices predicted between 5 and 10 residues after an N-terminal methionine. For PHDpsihtm08, this reduced the falsely predicted signal peptides from 322 (23%) to 146 (10%). Encouragingly, when we applied the same rule to the set of membrane proteins, no helix was removed by this rule. For three out of the 1418 signal peptides, PHDpsihtm08 incorrectly predicted two transmembrane helices.

Discussion

Confirming previous analyses

Some methods correctly distinguish globular from helical membrane proteins. Previous analyses showed that simple hydrophobicity-based methods have problems distinguishing between helical transmembrane and globular proteins (Edelman 1993; Jones et al. 1994; Rost et al. 1995; Jayasinghe et al. 2001a; Möller et al. 2001). In general, we confirmed this finding (Table 5). However, the Wolfenden and the Ben-Tal scales were clearly exceptional in this respect. Both performed on a par with the best advanced methods that predict membrane helices in at most 3% of all globular proteins (Table 5). Interestingly, these levels of accuracy are similar to the performance of the same methods six years

Table 6. *Incorrectly predicted membrane helices in signal peptides (false positives)*

Method	Percentage of proteins with signal peptides
ERROR	±1
PHDpsihtm08	23
PHDhtm08	24
Wolfenden	*26*
TMHMM1	*34*
PHDhtm07	*45*
PRED-TMR	*41*
HMMTOP2	*48*
Ben-Tal	*57*
SOSUI	*61*
TopPred2	*82*
WW	*90*
DAS	*97*
GES	*98*
Eisenberg	*99*
KD	*99*
Sweet	*99*
Hopp-Woods	*99*
Nakashima	*99*
Heijne	*99*
Levitt	*99*
Roseman	*99*
A-Cid	*99*
Av-Cid	*99*
Lawson	*99*
EM	*99*
Fauchere	*99*
Bull-Breese	*99*
Radzicka	*99*

Data set: 1418 sequence unique signal peptides from http://www.cbs.dtu.dk/ftp/signalp/ collected by Nielsen and colleagues (Nielsen et al. 1996, 1997a,b).
Numbers in italics: two standard deviations below the lowest false-positive rate.

ago (Rost et al. 1996a,b). This finding confirms that the globular proteins added to PDB over the last decade are not radically different from the structures that we knew before (Rost and Sander 1993; Rost 2001). Möller and colleagues published significantly more pessimistic estimates for the confusion between globular and membrane proteins (Möller et al. 2001). Whereas our estimates were based entirely on proteins of known structure, those from Möller et al. were based on proteins of unknown structure. Thus, we see two possible reasons for the difference between the two estimates. (1) Proteins in PDB differ from proteins in SWISS-PROT in their average length by almost a factor of 2 because structural biologists often have to truncate the proteins to obtain high-resolution structures. We might argue that the truncated regions are more likely to be confused with membrane helices than the regions for which structure is determined. (2) Many of the proteins used by Möller and colleagues may, in fact, contain membrane helices or signal peptides (for which the error is higher, Table 6). We suspect that the truth lies somewhere between the two extremes. Hence, our estimates for the confusion between globular and membrane proteins may be slightly optimistic.

Most methods confuse signal peptides and membrane helices. Möller et al. tested prediction methods on 34 signal and target peptides. They found that most methods incorrectly predicted these regions to contain membrane helices. We tested all 27 methods on 1418 sequence-unique signal peptides. Our results confirmed the previously uncovered trends (Table 6). However, the larger set that we used revealed that TMHMM1, which is one of the best methods in this respect, confuses >30% of the signal peptides with membrane helices rather than <10% as previously estimated (Möller et al. 2001). Most simple methods based only on hydrophobicity scales confused >90% of all the signal peptides with membrane helices (exception: Wolfenden scale, Table 6). The good news was that the error could be reduced by experts who discard all membrane helices predicted closer than 10 residues to an N-terminal methionine. In this best-case scenario, PHDhtm and PHDpsihtm falsely predicted only ~10% of the signal peptides as membrane helices. Possibly, combinations of membrane-optimized and signal-peptide-optimized programs could reduce this error rate.

Most methods identify most membrane helices. We confirmed (Ikeda et al. 2001; Jayasinghe et al. 2001b; Möller et al. 2001) that many methods correctly predict most membrane helices (Fig. 2). We also found the most common mistake to be the under- or overprediction of a single transmembrane helix. However, our results differed in detail from previous analyses (see below).

Resolving differences in previous analyses

Some methods are better; none is clearly best. Evaluations of membrane prediction methods are sometimes based on different definitions for performance accuracy. A particular example of the latter is to count a prediction of one long helix as correct although it stretches over two observed helices and thus misses the break in between the two. Another misleading standard procedure is to only report values covering one side of the coin, that is, only the values of correctly predicted as percentage of observed or vice versa. Here, we carefully evaluated all methods on identical data sets and compiled all reasonable scores for prediction accuracy. To simplify the complexity, we focused in our report on a relatively limited number of scores. Another problem with many previous analyses is that investigators have not estimated the error associated with a particular score. For example, from Table 1 we may conclude that HMMTOP2 is much better than TopPred2 when applying any measure for prediction accuracy. Although the numbers differed greatly, a thorough bootstrap experiment revealed that the performance of the two methods was indeed indis-

tinguishable. We compared the methods in a pairwise manner for each score of the high-resolution data set (Fig. 1). Some methods appeared more accurate than others. However, no method(s) performed consistently better than all others by more than one standard error (Fig. 1). Our estimates of error margins explained the numerical differences found between three analyses (Ikeda et al. 2001; Jayasinghe et al. 2001b; Möller et al. 2001).

Simple hydrophobicity-based methods less accurate than advanced methods. Möller et al. (2001) suggested that simple hydrophobicity scale-based methods predict membrane helices almost as accurately as the best advanced methods. We could not confirm this proposition. In contrast, we found that the best advanced methods were significantly more accurate than the best hydrophobicity-scale based methods, both in terms of per-segment and per-residue accuracy (Tables 2 and 3). The only possible exception may be the per-residue performance of the Ben-Tal scale for the low-resolution data (Table 3). However, we did confirm that, because of overprediction, a few hydrophobicity-scale-based methods identify the observed membrane helices at a level of accuracy similar to that of advanced methods in $Q_{\text{htm}}^{\%\text{obs}}$ in Tables 2 and 3. Jayasinghe et al. found that the WW hydrophobicity scale-based method that they introduced outperformed even the best advanced methods ("We find that [the] WW scale ... identifies TM helices of membrane proteins with an accuracy greater than 99%"; Jayasinghe et al. 2001a). We could also not confirm this finding, no matter which definition of prediction accuracy we compared. Nevertheless, the major problem with simple hydrophobicity-based methods is their failure on globular proteins (Table 5) and signal peptides (Table 6). In fact, the error of hydrophobicity scales depends on the length of the protein. For example, the high-resolution chains had an average length of ~215 residues, whereas low-resolution proteins were, on average, ~420 residues long. Although hydrophobicity scales correctly predicted all helices in 28%–65% of the short proteins (Table 2), they only detected 5%–29% for the long proteins (Table 3). In particular, the scale that performed best on the high-resolution set (KD) dropped in accuracy from 65% (high) to 13% (low), whereas the scale that performed most poorly on the short proteins in the high-resolution data (Wolfenden) became best for the long proteins in the low-resolution data. The Wolfenden scale also performed relatively well on globular proteins (Table 5) and on signal peptides (Table 6). The price for the lack of overprediction is a low accuracy in detecting membrane helices (underprediction). Overall, the most successful hydrophobicity scale appeared to be the Ben-Tal scale, which is based on the free energy of transferring an amino acid from water into the center of the hydrocarbon region of a lipid bilayer (Kessel and Ben-Tal 2002). It out-performed the Wolfenden scale for membrane proteins and for globular proteins, and it bested all other scales for the low-resolution set. Simple hydrophobicity scales obviously have tremendous importance for sequence analysis. However, to use them as the only criterion to predict membrane helices appears to be a bad idea.

Incorrect ranking by per-segment accuracy depends on definition of score. As discussed above, any attempt to rank prediction methods should account for the standard error in the estimated level of accuracy. A particular illustration of this finding is that different definitions of the accuracy in correctly predicting all helices (eq. 4) would slightly alter the ranks. For example, DAS scored worst among all advanced methods when an overlap of at least nine residues was required to consider a helix correctly predicted (definition introduced by Möller et al. 2001), but it appeared to be the third-best of all advanced methods when we applied the definition introduced by Ikeda et al. (2001) (see Supplementary Table 1; available online at http://www.proteinscience.org). When giving different ranks only for significant differences, this apparent contradiction was resolved. Most averages were relatively insensitive to whether we required an overlap of 3 or 9 residues between predicted and observed helix ($Q_{\text{ok}}3$ and $Q_{\text{ok}}9$ in Supplementary Table 1; available online at http://www.proteinscience.org). However, contrary to what has been claimed previously, some methods had lower averages when requiring nine overlapping residues. Similarly, for most methods the average scores did not change considerably when using the definition of Ikeda et al (Q_{ok}11Centre in Supplementary Table 1; available online at http://www.proteinscience.org). However, although the score was lower for most methods for which it differed from the other two, for a few it was actually higher. These were methods that tended to underpredict helices. Overall, the dependence of ranking on the definition of the score used underscored the need to standardize evaluations.

Similar prediction accuracy for prokaryotic and eukaryotic membrane proteins. Ikeda et al. (2001) found that prediction methods are consistently worse at predicting membrane proteins from eukaryotes than those from prokaryotes. We could not verify this finding. Both for the high- and for the low-resolution data sets, we found that some methods reached slightly higher levels on one than on the other (Table 4). However, the differences were not significant.

Novel findings

Low-resolution experiments not much more accurate than prediction methods. The low-resolution experiments differed substantially in their assignments of membrane helices from high-resolution experiments. In fact, for a small subset of 13 high-resolution chains, many prediction methods appeared to be as correct—or as incorrect—as previously deposited low-resolution experiments (Table 1). This problem

was also reflected in the substantial differences between the numerical scores for some of the methods. For example, DAS, TopPred2, and the PHDhtm series used partial information about 9 of the 36 high-resolution chains for development. For these methods, the scores on the 27 cross-validated high-resolution chains were similar to those for the 36 high-resolution chains (data not shown). However, the per-segment scores for the low-resolution sets differed from those for the high-resolution sets (Tables 2 and 3, in particular Q_{ok}). There are two possible explanations for this: either the low-resolution set contains new motifs, or the low-resolution experiments over- or underassign many helices. Such errors could result in a particularly poor performance in terms of predicting all TM helices correctly. In fact, for the set of 13 proteins for which we had low- and high-resolution experiments, Q_{ok} was low (84%, Table 1) for the low-resolution experiments. Furthermore, the observation that DAS, TopPred2, and the PHDhtm series got higher per-residue scores on the low-resolution data than on the high-resolution data indicated that the low-resolution assignments might not reflect completely new membrane motifs. Thus, the estimate for these cross-validated methods may be correctly estimated by the high-resolution data set (Table 2).

Problems with topology assignments by low-resolution data. The topologies of two proteins were incorrectly assigned by the low-resolution experiments (Table 1). These two proteins were (1) PDB: 1EHK:B/SWISS-PROT: COX2_THETH; and (2) PDB: 1EUL:A/SWISS-PROT: ATA2_RABIT. (1) 1EHK:B has one membrane helix and the N terminus is in the periplasm. Thus, PDB annotates the topology IN. In contrast, SWISS-PROT (release 34) annotates COX2_THETH with topology OUT, despite experimental data indicating otherwise (Keightley et al. 1995). Note that the latest SWISS-PROT release still annotates COX2_THETH as OUT. (2) The second pair is more complicated: The old SWISS-PROT release 20 entry for ATCA_RABIT was annotated with 10 membrane helices with topology IN, whereas the PDB structure 1EUL:A has 10 membrane helices with topology OUT. In contrast, the latest SWISS-PROT release for ATA2_RABIT annotates 10 helices, but still assigns the topology as IN according to antibody studies (Moller et al. 1997). However, this experimentally determined topology may be incorrect because of nonspecific antibodies for the N-terminus epitope. Indeed, the experimentalists noted that the antibody against the N terminus was only immunoreactive to the 1–243 N-terminal fragment rather than specific to the N-terminal 12 residues. At the same time, they argued that this antiserum can correctly locate the epitope for residues 1–12 (Juul et al. 1995). They suggested that the N terminus is cytoplasmic, but for other cytosolic loops, the authors observed enhanced antibody reactivities. Additionally, the N terminus may be OUT because after solubilization with $C_{12}E_6$, proteolysis did not drastically increase reactivity of antiserum 1–12. Furthermore, antisera to epitopes on all loop regions of ATA2_RABIT were not tested. Therefore, it would be useful to acquire information of the location of the other loops in ATA2_RABIT to verify the topological orientation of this protein.

All prediction methods missed only helices with weak experimental evidence. None of the helices in the high-resolution set and only three in the low-resolution set were missed by all advanced methods. As described above (in Results), the experiments done for these three proteins were not fully convincing in terms of the assignments of transmembrane helices and topology. This observation suggests implementing a consensus prediction of membrane helices. The potential success of such an approach has been initially tried out by a couple of authors (Promponas et al. 1999; Ikeda et al. 2001). However, these two initial attempts have focused only on advanced methods. Although advanced methods are more accurate than simple hydrophobicity-based methods, they tend to underpredict transmembrane helices, especially for high-resolution structures (Table 2). Advanced methods could thus serve as a specificity filter for a consensus method. Using both advanced and simple methods could help to verify low-resolution experimental results from proteolysis and gene fusion.

Not all membrane proteins identified. The only advanced method that predicted all known helical membrane proteins to contain at least one helix was DAS (Table 5, false negatives). However, the flip-side of the same coin was that DAS also performed poorly on globular proteins (Table 5, false positives). The other extreme was PHDhtm, based on conventional pairwise alignments that performed well in rejecting globular proteins while also missing almost one-fifth of the membrane proteins with the default parameters. Obviously, there is a tradeoff between predicting too many globular as membrane proteins, and too many membrane as globular proteins. Possibly the best compromise was achieved by SOSUI and TMHMM, which missed 6% of the membrane proteins while incorrectly predicting membrane helices in ~1% of all globular proteins. PHDhtm based on PSI-BLAST profiles (PHDpsihtm) reached a similar compromise: 8% of the membrane proteins were missed, and 2% of all globular proteins were mispredicted. Nevertheless, the problem of missing membrane proteins underlines once again that we need better methods that correctly distinguish between globular and membrane proteins.

Dependence of prediction accuracy on number of helices. We did not find any significant difference in the performance between proteins with one and many membrane proteins. In contrast, proteins with ≤ 5 membrane helices (≤ 5) were predicted more accurately than proteins with more (>6, Fig. 2B). Although we could label the difference as significant, we failed to come up with any reasonable explanation for this finding. Readers may speculate that the

numerical differences we observe between 6TM and 7TM proteins could be explained by the overabundance of transporters with buried charged residues. However, the number of proteins in each category was too small to validate such a fine-grained distinction.

Conclusion

We also overestimated the performance. Although we spent considerable effort on comparing prediction methods, our comparisons suffered from one crucial problem: We do not have cross-validation data available for all methods. In fact, the only methods for which we had cross-validated results were DAS, PHDhtm, PHDpsihtm, TopPred2, and most of the simple methods using only hydrophobicity scales. Although the overall scores for the advanced methods did not differ substantially between the sets of 27 cross-validated and 36-non-cross-validated high-resolution chains (data not shown), they did differ markedly between the nine chains used for development and the 27 cross-validated chains. This seemingly contradictory result is explained by the simple fact that most high-resolution proteins were not used in the development of these methods. In contrast, the newer prediction methods PRED-TMR, SOSUI, TMHMM, and WW used most and HMMTOP2 used all of the high-resolution chains for development. In fact, we observed two trends: (1) Newer methods were slightly better than older ones (HMMTOP2 was clearly more accurate than HMMTOP1 when tested on a small subset of the data); and (2) methods based on alignments were superior to those based on single sequences; in fact, when switching from using MaxHom (dynamic programming algorithm for conservation weight-based multiple sequence alignment) alignments against SWISS-PROT as input to PHDhtm to using PSI-BLAST alignments against all known sequences (BIG—nonidentical merger of SWISS-PROT and TrEMBL and PDB—and PHDpsiHtm), prediction accuracy increased considerably.

Most methods get most membrane helices, but the type of membrane protein is often wrong. The most common mistake was the under- or overprediction of one transmembrane helix. This appears encouraging in terms of prediction methods, in general. However, membrane predictions are very important in the context of analyzing entire proteomes because the number and orientation of the helices typically reveal aspects about function. In fact, only the very best methods predict all helices and the topology more often correctly than not. We may rightfully argue that present methods are still not good enough. Because both the number of helices and their orientation can easily be altered by engineering (Nilsson and von Heijne 1998; Ota et al. 1998; Monne et al. 1999a,b), the task at hand is, however, not an easy one. These experiments along with our analysis of the conservation of transmembrane helices strongly argue against the view that the number and orientation of membrane helices constitute a "solid reality written into the sequence." Rather, single residue exchanges can alter these macroscopic features. Thus, correct predictions require a precision typically not achieved. Perhaps present methods have reached the maximum possible level of accuracy and the chapter of simply predicting the location and orientation of membrane proteins is closed. With the recent high-resolution structures challenging common assumptions and our present analysis highlighting the number of urgent problems with prediction methods, we strongly doubt this. Therefore, we challenge that the issues elucidated in this investigation have reopened the field rather than closed it.

Materials and methods

Data sets

High-resolution data sets for membrane proteins. We started with a total set of 105 chains from helical membrane proteins for which a high-resolution structure was deposited in PDB (Berman et al. 2000). We identified these as helical membrane proteins according to the excellent up-to-date collection of membrane proteins at http://blanco.biomol.uci.edu (Jayasinghe et al. 2001b).

Low-resolution data sets for membrane proteins. We used an expert-curated set of 165 helical membrane proteins that was collected by Stefan Möller and colleagues (Möller et al. 2000). For all these proteins, good low-resolution experimental evidence about localization was available. For the comparison between high-resolution and low-resolution data, we used the annotations we found about transmembrane helix location in old SWISS-PROT versions released prior to the publication of the high-resolution structures.

High-resolution data set for globular proteins. The EVA server (Eyrich et al. 2001) continuously maintains a sequence-unique subset of PDB proteins. We used the version from July 2001 with 1852 representative protein chains. From that set we first removed all membrane proteins. Then we removed all proteins that were similar to one representative in a SCOP superfamily (Murzin et al. 1995; Lo Conte et al. 2000). Representatives were taken to be the longest proteins in the respective superfamily. This procedure yielded a final set of 616 globular protein chains.

Data set of proteins with known signal peptides. Henrik Nielsen and colleagues at the CBS in Copenhagen keep an up-to-date list of experimentally known signal peptides at their Web site (http://www.cbs.dtu.dk/ftp/signalp/readme). This group also spent considerable effort at defining thresholds for what constitutes redundancy in sets of signal peptides (Nielsen et al. 1996, 1997a). We downloaded a set of 1418 sequence-unique signal peptides from a total list of 2845.

Sequence-unique subsets reduce bias. Many of the proteins for which we have information about TM regions are similar to one another. If we want to analyze prediction methods or simple features such as TM length, this bias is problematic. To reduce the bias from the set of enzymes of known function, we have to first generate all-against-all alignments that capture the bias existing in that set. Then, we have to choose the maximal subset that fulfils the constraint that no pair in that subset is sequence-similar. Technically, we accomplished this objective in the following way. First, a pairwise BLAST (Altschul and Gish 1996) aligned all membrane proteins against each other. Second, the resulting pairs were filtered applying the HSSP-threshold (value $\theta = 0$, below) such that all remaining pairs were likely to have similar structures. Third,

the resulting families were sorted by number of members and length. Fourth, all pairs were clustered with a simple greedy algorithm starting with the largest and longest families (Hobohm et al. 1992). Note that the threshold chosen roughly translated to "no pair with more than 33% sequence identity over more than 100 residues aligned." In particular, we used the following formula to compile the distance DIST from the HSSP-curve HSSP_PIDE (Rost 1999):

$$\text{DIST} = \text{PIDE} - \text{HSSP\_PIDE}(\vartheta)$$

$$\text{HSSP\_PIDE}(\vartheta) = \vartheta + \begin{cases} 100 & \text{, for } L \geq 11 \\ 480 \cdot L^{-0.32 \cdot (1 + e^{-L/1000})} & \text{, for } L \leq 450 \\ 19.5 & \text{, for } L > 450 \end{cases} \quad (1)$$

where PIDE is the percentage pairwise sequence identity (ignoring gaps and insertions). This procedure yielded 36 proteins in the high-resolution set, and 165 proteins in the low-resolution set.

Programs tested

Building multiple alignments. Two different alignment schemes were explored: (1) the dynamic programming method MaxHom (Sander and Schneider 1991), and (2) a profile-based PSI-BLAST (Altschul et al. 1997). The particular protocol for finding similarities with PSI-BLAST applied the usual precautions to avoid drift and pollution (Jones 1999; Przybylski and Rost 2002). Searches were restricted to three iterations, and the iteration parameter (*H*-value) to 10^{-10} was set. The search databases were SWISS-PROT (Bairoch and Apweiler 2000) and BIG (SWISS-PROT [Bairoch and Apweiler 2000] + TrEMBL [Bairoch and Apweiler 2000] + PDB [Berman et al. 2000]). To explore the conservation of membrane helices, we filtered all MaxHom alignments according to various distances θ (eq. 1).

Advanced prediction methods. We referred to prediction methods as advanced when they implement more than simple hydrophobicity scales. We tested the following programs: DAS, HMMTOP (version 2), PHDhtm, PHDpsihtm, PRED-TMR, SOSUI, TMHMM (version 2), and TopPred2. TopPred2 averages the GES-scale of hydrophobicity (Engelman et al. 1986) using a trapezoid window (von Heijne 1992; Sipos and von Heijne 1993). PHDhtm combines a neural network using evolutionary information with a dynamic programming optimization of the final prediction (Rost et al. 1995, 1996b). DAS optimizes the use of hydrophobicity plots (Cserzö et al. 1997). SOSUI (Hirokawa et al. 1998) uses a combination of hydrophobicity and amphiphilicity preferences to predict membrane helices. TMHMM is the most advanced, and seemingly most accurate, present method to predict membrane helices (Sonnhammer et al. 1998). It embeds a number of statistical preferences and rules into a hidden Markov model to optimize the prediction of the localization of membrane helices and their orientation (note: similar concepts are used for HMMTOP; Tusnady and Simon 1998). PRED-TMR uses a standard hydrophobicity analysis with emphasis on detecting the ends and beginnings of membrane helices (Pasquier et al. 1999).

Simple methods exclusively based on hydrophobicity scales. We also implemented our in-house prediction methods that simply used various hydrophobicity scales for prediction. In particular, we tested the following scales: A-Cid, normalized hydrophobicity scale for α-proteins (Cid et al. 1992); Av-Cid, normalized average hydrophobicity scale (Cid et al. 1992); Ben-Tal, Hydrophobicity scale representing free energy of transfer of an amino acid from water into the center of the hydrocarbon region of a model lipid bilayer (Kessel and Ben-Tal 2002); Bull-Breese, Bull-Breese hydrophobicity scale (Bull 1974); Eisenberg, normalized consensus hydrophobicity scale (Eisenberg et al. 1984); EM, Solvation free energy (Eisenberg and McLachlan 1986); Fauchere, hydrophobic parameter π from the partitioning of *N*-acetyl-amino-acid amides (Fauchere and Pliska 1983); GES, hydrophobicity property (Engelman et al. 1986; Prabhakaran 1990); Heijne, transfer free energy to lipophilic phase (von Heijne and Blomberg 1979); Hopp-Woods, Hopp-Woods hydrophilicity value (Hopp and Woods 1981); KD, Kyte–Doolittle hydropathy index (Kyte and Doolittle 1982); Lawson, transfer free energy (Lawson et al. 1984); Levitt, hydrophobic parameter (Levitt 1976); Nakashima, normalized composition of membrane proteins (Nakashima et al. 1990); Radzicka, transfer free energy from 1-octanol to water (Radzicka and Wolfenden 1988); Roseman, solvation-corrected side-chain hydropathy (Roseman 1988); Sweet, optimal matching hydrophobicity (Sweet and Eisenberg 1983); Wolfenden, hydration potential (Wolfenden et al. 1981); and WW, Wimley–White scale (Jayasinghe et al. 2001a). Replacing the WW scale with each of the above-mentioned hydrophobicity indices, we used the WW algorithm to evaluate the predictive performance of each index.

Measuring accuracy

Measuring per-segment accuracy. The ultimate goal of prediction methods obviously is to correctly predict all residues. Assume a protein with 10 membrane helices of 20 residues each; method A predicts 10 helices but gets the five residues at each end of each helix wrong, and method B misses four helices but gets the ends for the other six entirely right. Which method is better? Possibly, many readers would favor method A. This problem is captured in using two different scores measuring prediction accuracy in the field of globular secondary structure prediction: per-residue scores and per-segment scores (Rost and Sander 1993; Rost et al. 1994). Although globular secondary-structure segments are, on average, rather short (helices ~10 residues, strands ~5 residues), membrane helices are rather long. Consequently, the problem of evaluating the per-segment accuracy allows a more coarse-grained measure than required for globular secondary-structure prediction (Rost et al. 1994; Zemla et al. 1999). There are two separate issues to address when defining a helix to be predicted correctly. The first concerns counting the same helix twice. We used the simple concept of "correctly predicted segment" shown in Figure 4.

In particular, the observed helix O2 is not correctly predicted, because P1 overlaps already with O1. Similarly, P2 is counted as correct with respect to O3, whereas P3 is not. The second issue concerns the minimal overlap required between the observed and predicted helix. If not stated otherwise, we required a minimal overlap of 3 residues, following the definitions previously used in many other publications (von Heijne 1992; Jones et al. 1994; Persson and Argos 1994; von Heijne 1994; Rost et al. 1995, 1996b;

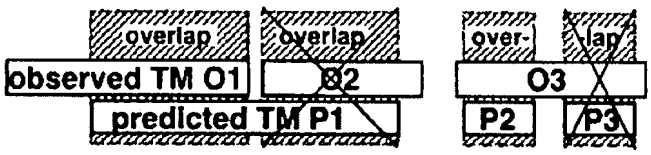


Fig. 4. Correctly predicted segments. In this example, there are three observed and three predicted helices. Observed helix O1 is correctly predicted by P1 as they overlap. However, observed helix O2 is not correctly predicted because P1 already overlaps with O1. Hence, P1 cannot be used as a correct prediction for O2. Similarly, P2 is counted as correct only with respect to O3, whereas P3 is not since O3 was already predicted by P2.

Persson and Argos 1996; Sonnhammer et al. 1998). Möller et al. (2001) used a similar procedure; however, they required an overlap of at least 9 rather than 3 residues. Other groups required a minimal overlap of 1 residue (e.g., Cserzö et al. 1997; Tusnady and Simon 1998). Jayasinghe required an overlap of 9 (Jayasinghe et al. 2001b) and 3 (Jayasinghe et al. 2001a) residues; however, in both publications, they counted the same predicted helix twice, thus yielding 100% accuracy for the overlap between O1/P1 and O2/P2 in Figure 4. Yet another measure was introduced by Ikeda et al. (2001): Helices were considered as correctly predicted if the centers of the predicted and the observed helix overlapped by at least 11 residues. The different measures are illustrated in the following example for a prediction (T = transmembrane):

```
observed:
-TTTTTTTTTTTTTTTTTTTT——TTTTTTTTTTTTTTTTTTT-
predict 1: ————————-TTTTTTTTTT————————-
predict 2: ————————TTTTTTTTTTTTTTTT————-
predict 3: ————-TTTTTTTTTTTTTTTTTTTT————-
predict 4: ————TTTTTTTTTTTTTTTTTTTTT————
```

Jayasinghe et al. (2001a) evaluates prediction 1 as 0% accurate and 2–4 as 100% accurate (two helices correct); Jayasinghe et al. (2001b) give predictions 1 and 2 0% and 3 and 4 100%; Tusnady and Simon (1998) give 1–4 50% (one helix right, one not); Möller et al. (2001) give 1–2 0% and 3–4 50%; Ikeda et al. (2001) give 1–3 0% and 4 50%; the score that we refer to in this manuscript gives 1 0% and 2–4 50%. For comparison, we also provided a few other scores in the Supplementary Material (available online at http://www.proteinscience.org; note that we, however, did not count helices twice in any of those definitions).

With this concept, we can compile the percentage of correctly predicted transmembrane helices:

$$Q_{\text{htm}}^{\%\text{obs}} = 100 \cdot \frac{\text{number of correctly predicted TM in data set}}{\text{number of TM observed in data set}} \tag{2}$$

where $Q_{\text{htm}}^{\%\text{obs}}$ estimates the likelihood that an actual membrane helix is correctly predicted. Although this score can also be compiled for a single protein, it would be misleading to compile the score for each protein in a data set and then to average over all proteins. Rather, the number should be compiled by pooling all membrane helices from an entire data set. Overpredictions are measured by the corresponding score:

$$Q_{\text{htm}}^{\%\text{prd}} = 100 \cdot \frac{\text{number of correctly predicted TM in data set}}{\text{number of TM predicted in data set}} \tag{3}$$

where $Q_{\text{htm}}^{\%\text{prd}}$ estimates the likelihood that a predicted TM is correctly predicted. These two scores are merged into a score that describes for which percentage of the proteins all TM segments are correctly predicted:

$$Q_{\text{ok}} = \frac{100}{N_{\text{prot}}} \cdot \sum \delta_i, \text{ with } \delta_i = \begin{cases} 1, \text{ if } Q_{\text{htm}}^{\%\text{obs}} \wedge Q_{\text{htm}}^{\%\text{prd}} = 100 \\ 0, \text{ else} \end{cases} \tag{4}$$

Thus, Q_{ok} becomes 100 if and only if for all proteins in the set both $Q_{\text{htm}}^{\%\text{obs}}$ and $Q_{\text{htm}}^{\%\text{prd}}$ reach 100%. Finally, we need to evaluate the accuracy of predicting the topology correctly:

$$\text{TOPO} = 100 \cdot \frac{\text{number of proteins with correctly predicted topology}}{\text{number of proteins}}. \tag{5}$$

Measuring per-residue accuracy. Although the per-segment scores capture most of what experts would intuitively consider as important features of TMH prediction methods, we also need to monitor a number of per-residue scores that evaluate how accurately particular residues are predicted. In particular, the example of P2 and P3 in Figure 4 would yield 0 for all per-segment scores, although the predictions somehow capture important information. The simplest per-residue score is the two-state per-residue accuracy Q_2, which measures the percentage of residues predicted correctly in either of the two states T (membrane helix) or N (not membrane):

$$Q_2 = \frac{100}{N_{\text{prot}}} \cdot \sum_{i=1}^{N_{\text{prot}}} \frac{\text{number of residues predicted correctly in protein i}}{\text{number of residues in protein i}} \tag{6}$$

Typically, most residues in membrane proteins are in globular regions (Liu and Rost 2001). Thus, nonmembrane residues tend to dominate Q_2. This problem can be overcome by simply measuring the percentage of residues correctly predicted in membrane segments:

$$Q_{\text{sT}}^{\%\text{obs}} = 100 \cdot \frac{\text{number of residues correctly predicted in TM helices}}{\text{number of residues observed in TM helices}} \tag{7}$$

Similar to the per-segment scores, overpredictions can be captured by the corresponding score:

$$Q_{\text{2T}}^{\%\text{prd}} = 100 \cdot \frac{\text{number of residues correctly predicted in TM helices}}{\text{number of residues predicted in TM helices}} \tag{8}$$

$Q_{\text{2N}}^{\%\text{obs}}$ and $Q_{\text{2N}}^{\%\text{prd}}$ are the corresponding percentages for nonmembrane residues. Finally, we monitored the Matthews correlation index (Matthews 1975) that attempts to capture both over- and underprediction of residues in transmembrane helices by one single score. This index is defined as:

$$C_{\text{T}} = \frac{p_{\text{T}} \cdot n_{\text{T}} - u_{\text{T}} \cdot o_{\text{T}}}{\sqrt{(p_{\text{T}} + u_{\text{T}}) \cdot (p_{\text{T}} + o_{\text{T}}) \cdot (n_{\text{T}} + u_{\text{T}}) \cdot (n_{\text{T}} + o_{\text{T}})}} \tag{9}$$

where p_{T} is the number of residues correctly predicted as membrane helix (TMH), n_{T} is the number of residues correctly predicted as non-TMH, and u_{T} and o_{T} are the number of residues under- and overpredicted, respectively.

Estimating error for per-residue accuracy: standard error. For globular proteins, prediction accuracy varies considerably between different proteins (Rost et al. 1993; Rost 1996). The corresponding distributions can be approximated by Gaussian distributions. Thus, we can estimate the standard error of score Q by the simple rule-of-thumb:

$$\text{SE}(Q) = \frac{\sigma\{Q{:}N_{\text{prot-large}}\}}{\sqrt{N_{\text{prot-set}}}} \tag{10}$$

where σ is the standard deviation for score Q based on a data set of $N_{\text{prot-large}}$ proteins. This set has to be sufficiently large to actually observe a normal distribution. Assuming that we only have a much smaller data set of $N_{\text{prot-set}}$ proteins, we can then still approximate the standard error by using the standard deviation compiled over the large data set. Whereas this concept is easy to apply to evaluations of globular prediction methods (Eyrich et al. 2001; Rost and Eyrich 2001), for the situation of membrane proteins, we simply do not have a sufficient number of high-resolution structures to once and for all estimate σ. There is no clean solution to this problem. Here, we used the following approximation:

$$SE(Q) \leftarrow 1/\sqrt{N_{protX}} \cdot \max_{\left\{\substack{\text{all methods for set X;}\\ \text{all sets Y with } N_{protY} \geq N_{protX}\}\right.} \sigma \qquad (11)$$

that is, we used the maximal possible standard error. Assume that $\sigma = 20$ for a set of 13 proteins, $\sigma = 10$ for a set of 36 proteins, and $\sigma = 15$ for a set of 27 proteins. Then we used $\sigma = 20$ for the first, and $\sigma = 15$ for the other two.

Estimating error for per-segment accuracy: bootstrap experiment. The above concept to estimate the error in evaluating performance is not applicable for the per-segment scores, because these are not distributed normally. To illustrate the problem for the topology prediction: scores can be 1 (correct topology) or 0 (incorrect) for one protein. The score TOPO (eq. 5) averages over all proteins, hence provides one single final value, rather than a distribution. One way to still estimate the error in such a situation is the bootstrap experiment (Diaconis and Efron 1983; Efron et al. 1996). The procedure is the following (Fig. 5): (1) Assume we have a set of $N = 36$ proteins, each with correct or incorrect topology. (2) Choose a random subset of $K < N$ proteins, and compile the average (TOPO) over these K proteins. (3) Repeat M times and estimate the error based on the resulting distribution of averages. In other words, the bootstrapping experiment attempts to estimate how sensitively a score depends on a particular data set chosen. Albeit often surprisingly powerful, bootstrapping is a more coarse-grained approximation. In particular, we used the following parameters to estimate errors for per-segment scores: $M = 100$ (100 random picks), and $K = \text{int}(N/2)$; that is, for each random pick we chose half of the proteins available in the respective sets. Finally, we applied the same approximation as depicted in equation 11, that is, reported a rather conservative estimate for the error.

Ranking methods. Given methods A and B evaluated on a set with N proteins, when can we conclude that the performance of A (Q(A)) is significantly better than that of B (Q(B))? The error estimates provide an answer to this question: We cannot distinguish between A and B if:

$$\Delta Q = Q(A) \leftarrow Q(B) \leq SE(Q) \qquad (12)$$

Thus, we can rank only if A and B differ by more than the error. For example, when a method correctly predicts 75% of the residues in a test set of 16 proteins with a standard deviation of 10%, a difference relative to another method that is smaller than 2.5% (i.e., $\Delta Q = 10/\text{sqrt}[16]$) is not significant. Thus, we cannot distinguish between two methods that predict correctly 75% and 73% of all residues, respectively. We used this estimate to rank methods in the following way. Assume four methods have accuracy levels of A = 75, B = 73, C = 71, and D = 68. D can be distinguished from all other methods ($\Delta Q > 2.5$ to all). Hence, it ranks last. C can be distinguished from A ($\Delta Q = 4 > 2.5$). However, A cannot be distinguished from B ($\Delta Q = 2 < 2.5$), and B cannot be distinguished from C ($\Delta Q = 2 < 2.5$). This situation results in a dilemma that has four different possible solutions: (I) A, B, and C get the same rank, ascertaining that no two methods are ranked differently that cannot be distinguished. (II) A and B get rank 1, and C rank 2, ensuring that no two methods are ranked equally that can be distinguished. (III) A gets rank 1, B rank 2, and C rank 3, ignoring that we cannot distinguish between A and B, nor between B and C. (IV) Do not rank. None of these solutions is correct. Here, we applied solutions (IV) and (I). For the example given, solution (I) implied that A, B, and C ranked first; D ranked second. However, this simplification ignored another intrinsically insurmountable problem: What if method A is significantly better than method B in terms of Q_2 and significantly worse in terms of Q_{ok}? Occasionally, the following ad hoc solution is presented to such a problem: Rank all methods on all scores and compile averages over ranks (Tables 3 and 5).

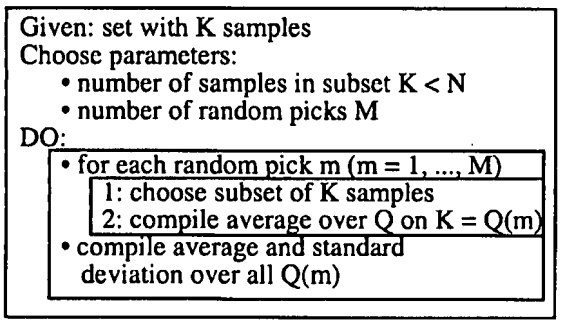


Fig. 5. Procedure for estimating error using a bootstrap experiment. Given a data set with N items, one first defines K, which is the number of items one will select from the original data set, and M, which is the number of times one will choose a sample of size K. For instance, if the data set is of size 36, then one defines $K < 36$. Once K and M are defined, one selects a sample of size K and calculates the average value for the appropriate metric. Repeating this process M times will yield M average values. One can then compile the averaged value and standard deviation for these M average values.

Electronic supplemental material

All data sets and a few additional results are available through our Web site at: http://cubic.bioc.columbia.edu/papers/2002_htm_eval/data.

Acknowledgments

Thanks to Jinfeng Liu (Columbia) for computer assistance and the collection of genome data sets; and to Jinfeng Liu and Dariusz Przybylski (Columbia) for providing preliminary information and programs. Particular thanks to Volker Eyrich (Columbia) for making the META-PredictProtein server available! The work of B.R. was supported by the grants 1-P50-GM62413-01 and RO1-GM63029-01 from the National Institutes of Health. Last, but not least, thanks to all those who deposit their experimental data in public databases, and to those who maintain these databases.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Altschul, S.F. and Gish, W. 1996. Local alignment statistics. *Meth. Enzymol.* **266:** 460–480.

Altschul, S., Madden, T., Shaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Amstutz, P., Forrer, P., Zahnd, C., and Pluckthun, A. 2001. In vitro display technologies: Novel developments and applications. *Curr. Opin. Biotechnol.* **12:** 400–405.

Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28:** 45–48.

Bauer, M.F., Hofmann, S., Neupert, W., and Brunner, M. 2000. Protein translocation into mitochondria: The role of TIM complexes. *TICB* **10:** 25–31.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28:** 235–242.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* **112:** 535–542.

Bettler, B., Kaupmann, K., and Bowery, N. 1998. GABAB receptors: Drugs meet clones. *Curr. Opin. Neurobiol.* **8:** 345–350.

Bull, H.B.a.B. 1974. Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch. Biochem. Biophys.* **161:** 665–670.

Chapman, R., Sidrauski, C., and Walter, P. 1998. Intracellular signaling from the endoplasmic reticulum to the nucleus. *Annu. Rev. Cell Dev. Biol.* **14:** 459–485.

Chen, X. and Schnell, D.J. 1999. Protein import into chloroplasts. *Trends Cell Biol.* **9:** 222–227.

Cid, H., Bunster, M., Canales, M., and Gazitua, F. 1992. Hydrophobicity and structural classes in proteins. *Prot. Engin.* **5:** 373–375.

Cserzö, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. 1997. Prediction of transmembrane α-helices in prokaryotic membrane proteins: The dense alignment surface method. *Prot. Engin.* **10:** 673–676.

Diaconis, P. and Efron, B. 1983. Computer-intensive methods in statistics. *Scient. Am.* **248:** 96–108.

Edelman, J. 1993. Quadratic minimization of predictors for protein secondary structure: Application to transmembrane α-helices. *J. Mol. Biol.* **232:** 165–191.

Efron, B., Halloran, E., and Holmes, S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci.* **93:** 13429–13434.

Eisenberg, D. and McLachlan, A.D. 1986. Solvation energy in protein folding and binding. *Nature* **319:** 199–203.

Eisenberg, D., Weiss, R.M., and Terwilliger, T.C. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci.* **81:** 140–144.

Engelman, D.M., Steitz, T.A., and Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15:** 321–353.

Eyrich, V. and Rost, B. 2000. The META-PredictProtein server. CUBIC, Columbia University, Department of Biochemistry and Molecular Biophysics. http://cubic.bioc.columbia.edu/predictprotein/submit_meta.html/.

Eyrich, V., Martí-Renom, M.A., Przybylski, D., Fiser, A., Pazos, F., Valencia, A., Sali, A., and Rost, B. 2001. EVA: Continuous automatic evaluation of protein structure prediction servers. *Bioinformatics* **17:** 1242–1243.

Fauchere, J.L. and Pliska, V. 1983. Hydrophobic parameters π of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18:** 369–375.

Gentschev, I. and Goebel, W. 1992. Topological and functional studies on HlyB of *Escherichia coli*. *Mol. Gen. Genet.* **232:** 40–48.

Hennessey, E.S. and Broome-Smith, J.K. 1993. Gene-fusion techniques for determining membrane-protein topology. *Curr. Opin. Str. Biol.* **3:** 524–531.

Hettema, E.H., Distel, B., and Tabak, H.F. 1999. Import of proteins into peroxisomes. *Biochim. Biophys. Acta* **1451:** 17–34.

Heusser, C. and Jardieu, P. 1997. Therapeutic potential of anti-IgE antibodies. *Curr. Opin. Immunol.* **9:** 805–813.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. 1998. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14:** 378–379.

Hobohm, U., Scharf, M., Schneider, R., and Sander, C. 1992. Selection of representative protein data sets. *Protein Sci.* **1:** 409–417.

Hofmann, K. and Stoffel, W. 1993. TMBASE—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374:** 166.

Hopp, T.P. and Woods, K.R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci.* **78:** 3824–3828.

Ikeda, M., Arai, M., Lao, D.M., and Shimizu, T. 2001. Transmembrane topology prediction methods: A reassessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol.* **1:** http://www.bioinfo.de/isb/2001/2002/0003/.

Ito, A. 2000. Mitochondrial processing peptidase: Multiple-site recognition of precursor proteins. *TICB* **10:** 25–31.

Jayasinghe, S., Hristova, K., and White, S.H. 2001a. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **312:** 927–934.

———. 2001b. MPtopo: A database of membrane protein topology. *Protein Sci.* **10:** 455–458.

Jermutus, L., Ryabova, L.A., and Pluckthun, A. 1998. Recent advances in producing and selecting functional proteins by using cell-free translation. *Curr. Opin. Biotechnol.* **9:** 534–548.

Johnson, J.M. and Church, G.M. 1999. Alignment and structure prediction of divergent protein families: Periplasmic and outer membrane proteins of bacterial efflux pumps. *J. Mol. Biol.* **287:** 695–715.

Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292:** 195–202.

Jones, D.T., Taylor, W.R., and Thornton, J.M. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochem.* **33:** 3038–3049.

Jording, D. and Puhler, A. 1993. The membrane topology of the *Rhizobium meliloti* C4-dicarboxylate permease (DctA) as derived from protein fusions with *Escherichia coli* K12 alkaline phosphatase (PhoA) and β-galactosidase (LacZ). *Mol. Gen. Genet.* **241:** 106–114.

Juretic, D., Zucic, D., Lucic, B., and Trinajstic, N. 1998. Preference functions for prediction of membrane-buried helices in integral membrane proteins. *Comput. Chem.* **22:** 279–294.

Juul, B., Turc, H., Durand, M., de Gracia, A., Denoroy, L., Moller, J., Champeil, P., and le Maire, M. 1995. Do transmembrane segments in proteolyzed sarcoplasmic reticulum Ca^{2+}-ATPase retain their functional Ca^{2+} binding properties after removal of cytoplasmic fragments by proteinase K? *J. Biol. Chem.* **270:** 20123–20134.

Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22:** 2577–2637.

Keightley, J., Zimmerman, B., Mather, M., Springer, P., Pastusztyn, A., Lawrence, D., and Fee, J. 1995. Molecular genetic and protein chemical characterization of the cytochrome ba3 from *Thermus thermophilus* HB8*. *J. Biol. Chem.* **270:** 20345–20358.

Kessel, A. and Ben-Tal, N. 2002. Free energy determinants of peptide association with lipid bilayers. In *Peptide–lipid interactions* (eds. S. Simon and T. McIntosh). Academic Press, San Diego (in press).

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305:** 567–580.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Lawson, E.Q., Sadler, A.J., Harmatz, D., Brandau, D.T., Micanovic, R., MacElroy, R.D., and Middaught, C.R. 1984. A simple experimental model for hydrophobic interactions in proteins. *J. Biol. Chem.* **259:** 2910–2912.

Le Borgne, R. and Hoflack, B. 1998. Protein transport from the secretory to the endocytic pathway in mammalian cells. *Biochim. Biophys. Acta* **1404:** 195–209.

Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104:** 59–107.

Lewis, M.J., Chang, J.A., and Simoni, R.D. 1990. A topological analysis of subunit A from *Escherichia coli* F_1F_0-ATP synthase predicts eight transmembrane segments. *J. Biol. Chem.* **265:** 10541–10550.

Liu, J. and Rost, B. 2001. Comparing function and structure between entire proteomes. *Protein Sci.* **10:** 1970–1979.

Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G., and Chothia, C. 2000. SCOP: A structural classification of proteins database. *Nucleic Acids Res.* **28:** 257–259.

Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. 2002. SCOP database in 2002: Refinements accommodate structural genomics. *Nucleic Acids Res.* **30:** 264–267.

Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B., and Sali, A. 2002. Reliability of assessment of protein structure prediction methods. *Structure* **10:** 435–440.

Matthews, B.W. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **405:** 442–451.

McGovern, K., Ehrmann, M., and Beckwith, J. 1991. Decoding signals for membrane proteins using alkaline phosphatase fusions. *EMBO J.* **10:** 2773–2782.

McGuigan, J.E. 1994. Antibodies to complementary peptides as probes for receptors. *Immunomethods* **5:** 158–166.

Miroux, B., Frossard, V., Raimbault, S., Ricquier, D., and Bouillaud, F. 1993. The topology of the brown adipose tissue mitochondrial uncoupling protein determined with antibodies against its antigenic sites revealed by a library of fusion proteins. *EMBO J.* **12:** 3739–3745.

Moller, J., Ning, G., Maunsebach, A., Fujimoto, K., Asai, K., Juul, B., Lee, T.-J., de Gracia, A., Falson, P., and le Maire, M. 1997. Probing of the membrane topology of sarcoplasmic reticulum Ca^{+2}-ATPase with sequence-specific antibodies. *J. Biol. Chem.* **272:** 29015–29032.

Möller, S., Kriventseva, E.V., and Apweiler, R. 2000. A collection of well characterised integral membrane proteins. *Bioinformatics* **16:** 1159–1160.

Möller, S., Croning, D.R., and Apweiler, R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17:** 646–653.

Monne, M., Gafvelin, G., Nilsson, R., and von Heijne, G. 1999a. N-Tail translocation in a eukaryotic polytopic membrane protein: Synergy between neighboring transmembrane segments. *Eur. J. Biochem.* **263:** 264–269.

Monne, M., Nilsson, I., Elofsson, A., and von Heijne, G. 1999b. Turns in transmembrane helices: Determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J. Mol. Biol.* **293:** 807–814.

Moreau, J.L. and Huber, G. 1999. Central adenosine A(2A) receptors: An overview. *Brain Res. Brain Res. Rev.* **31:** 65–82.

Morris, G.E., Sedgwick, S.G., Ellis, J.M., Pereboev, A., Chamberlain, J.S., and Nguyen thi, M. 1998. An epitope structure for the C-terminal domain of dystrophin and utrophin. *Biochemistry* **37:** 11117–11127.

Moult, J., Pedersen, J.T., Judson, R., and Fidelis, K. 1995. A large-scale experiment to assess protein structure prediction methods. *Proteins* **23:** ii–iv.

Moult, J., Hubbard, T., Bryant, S.H., Fidelis, K., and Pedersen, J.T. 1997. Critical assessment of methods of protein structure prediction (CASP): Round II. *Proteins* **Suppl 1:** 2–6.

Moult, J., Hubbard, T., Fidelis, K., Pedersen, J.T. 1999. Critical assessment of methods of protein structure prediction (CASP): Round III. *Proteins* **Suppl 3:** 2–6.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247:** 536–540.

Nakashima, H., Nishikawa, K., and Ooi, T. 1990. Distinct character in hydrophobicity of amino acid composition of mitochondrial proteins. *Proteins* **8:** 173–178.

Neuwald, A.F., Liu, J.S., and Lawrence, C.E. 1995. Gibbs motif sampling: Detection of bacterial outer membrane protein repeats. *Protein Sci.* **4:** 1618–1631.

Nielsen, H., Engelbrecht, J., von Heijne, G., and Brunak, S. 1996. Defining a similarity threshold for a functional protein sequence pattern: The signal peptide cleavage site. *Proteins* **24:** 165–177.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. 1997a. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot. Engin.* **10:** 1–6.

———. 1997b. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Intl. J. Neural Sys.* **8:** 581–599.

Nilsson, I. and von Heijne, G. 1998. Breaking the camel's back: Proline-induced turns in a model transmembrane helix. *J. Mol. Biol.* **284:** 1185–1189.

Ota, K., Sakaguchi, M., von Heijne, G., Hamasaki, N., and Mihara, K. 1998. Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. *Mol. Cell* **2:** 495–503.

Pahl, H.L. 1999. Signal transduction from the endoplasmic reticulum to the cell nucleus. *Physiol. Rev.* **79:** 683–701.

Pasquier, C., Promponas, V.J., Palaios, G.A., Hamodrakas, J.S., and Hamodrakas, S.J. 1999. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: The PRED-TMR algorithm. *Prot. Engin.* **12:** 381–385.

Persson, B. and Argos, P. 1994. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* **237:** 182–192.

———. 1996. Topology prediction of membrane proteins. *Protein Sci.* **5:** 363–371.

Prabhakaran, M. 1990. The distribution of physical, chemical and conformational properties in signal and nascent peptides. *Biochem. J.* **269:** 691–696.

Promponas, V.J., Palaios, G.A., Pasquier, C.M., Hamodrakas, J.S., and Hamodrakas, S.J. 1999. CoPreTHi: A Web tool which combines transmembrane protein segment prediction methods. *In Silico Biol.* **1:** 159–162.

Przybylski, D. and Rost, B. 2002. Alignments grow, secondary structure prediction improves. *Proteins* **46:** 195–205.

Radzicka, A. and Wolfenden, R. 1988. Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27:** 1664–1670.

Roseman, M.A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200:** 513–522.

Rost, B. 1996. PHD: Predicting one-dimensional protein structure by profile based neural networks. *Meth. Enzymol.* **266:** 525–539.

———. 1999. Twilight zone of protein sequence alignments. *Prot. Engin.* **12:** 85–94.

———. 2001. Protein secondary structure prediction continues to rise. *J. Struct. Biol.* **134:** 204–218.

———. 2002. Enzyme function less conserved than anticipated. *J. Mol. Biol.* **318:** 595–608.

Rost, B. and Eyrich, V. 2001. EVA: Large-scale analysis of secondary structure prediction. *Proteins* **45 Suppl 5:** S192–S199.

Rost, B. and Sander, C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232:** 584–599.

Rost, B., Sander, C., and Schneider, R. 1993. Progress in protein structure prediction? *TIBS* **18:** 120–123.

———. 1994. Redefining the goals of protein secondary structure prediction. *J. Mol. Biol.* **235:** 13–26.

Rost, B., Casadio, R., Fariselli, P., and Sander, C. 1995. Prediction of helical transmembrane segments at 95% accuracy. *Protein Sci.* **4:** 521–533.

Rost, B., Casadio, R., and Fariselli, P. 1996a. Refining neural network predictions for helical transmembrane proteins by dynamic programming. In *Fourth International Conference on Intelligent Systems for Molecular Biology* (eds. D. States et al.), pp. 192–200. AAAI Press, St. Louis, MO.

———. 1996b. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* **5:** 1704–1718.

Sander, C. and Schneider, R. 1991. Database of homology-derived structures and the structural meaning of sequence alignment. *Proteins* **9:** 56–68.

Saragovi, H.U. and Gehring, K. 2000. Development of pharmacological agents for targeting neurotrophins and their receptors. *Trends Pharmacol. Sci.* **21:** 93–98.

Sedlacek, H.H. 2000. Kinase inhibitors in cancer therapy: A look ahead. *Drugs* **59:** 435–476.

Sipos, L. and von Heijne, G. 1993. Predicting the topology of eukaryotic membrane proteins. *Eur. J. Biochem.* **213:** 1333–1340.

Soltys, B.J. and Gupta, R.S. 2000. Mitochondrial proteins at unexpected cellular locations: Export of proteins from mitochondria from an evolutionary perspective. *Intl. Rev. Cytol.* **194:** 133–196.

Sonnhammer, E.L.L., von Heijne, G., and Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. In *Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB98)* (eds. J. Glasgow et al.), pp. 175–182. AAAI Press, Montreal, Canada.

Stack, J.H., Horazdovsky, B., and Emr, S.D. 1995. Receptor-mediated protein sorting to the vacuole in yeast: Roles for a protein kinase, a lipid kinase and GTP-binding proteins. *Annu. Rev. Cell Dev. Biol.* **11:** 1–33.

Sweet, R.M. and Eisenberg, D. 1983. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. Mol. Biol.* **171:** 479–488.

Thanassi, D.G. and Hultlgren, S.J. 2000. Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr. Opin. Cell Biol.* **12:** 420–430.

Traxler, B., Boyd, D., and Beckwith, J. 1993. The topological analysis of integral membrane proteins. *J. Mem. Biol.* **132:** 1–11.

Truscott, K.N. and Pfanner, N. 1999. Import of carrier proteins into mitochondria. *Biol. Chem.* **380:** 1151–1156.

Tusnady, G.E. and Simon, I. 1998. Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. *J. Mol. Biol.* **283:** 489–506.

van Geest, M. and Lolkema, J.S. 2000. Membrane topology and insertion of membrane proteins: Search for topogenic signals. *Microbiol. Mol. Biol. Rev.* **64:** 13–33.

von Heijne, G. 1992. Membrane protein structure prediction. *J. Mol. Biol.* **225:** 487–494.

———. 1994. Membrane proteins: From sequence to structure. *Annu. Rev. Biophys. Biomol. Struct.* **23:** 167–192.

von Heijne, G. and Blomberg, C. 1979. Trans-membrane translocation of proteins: The direct transfer model. *Eur. J. Biochem.* **97:** 175–181.

Wallin, E. and von Heijne, G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7:** 1029–1038.

Wang, R., Seror, S., Blight, M., Pratt, J., Broome-Smith, J., and Holland, I. 1991. Analysis of the membrane organization of an *Escherichia coli* protein translocator, HlyB, a member of a large family of prokaryote and eukaryote surface transport proteins. *J. Mol. Biol.* **217:** 441–454.

White, S. 2001. Membrane proteins of known structure. http://blanco.biomol.uci.edu/mp_assembly.html/.

White, S.H. and Wimley, W.C. 1999. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28:** 319–365.

Wimley, W.C., Creamer, T.P., and White, S.H. 1996a. Solvation energies of amino acid side-chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35:** 5109–5124.

Wimley, W.C., Gawrisch, K., Creamer, T.P., and White, S.H. 1996b. A direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc. Natl. Acad. Sci.* **93:** 2985–2990.

Wolfenden, R.V., Cullis, P.M., and Southgate, C.C.F. 1979. Water, protein folding, and the genetic code. *Science* **206:** 575–577.

Wolfenden, R., Andersson, L., Cullis, P.M., and Southgate, C.C.B. 1981. Affinities of amino acid side chains for solvent water. *Biochemistry* **20:** 849–855.

Zemla, A., Venclovas, C., Fidelis, K., and Rost, B. 1999. A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment. *Proteins* **34:** 220–223.

CURRENT PROTOCOLS IN MOLECULAR BIOLOGY

VOLUME 2

John Wiley & Sons, Inc.

0-555-01748-6

CORE 20 (S66)

Library of Congress Cataloging in Publication Data:

Current protocols in molecular biology. 4 vols.

1. Molecular biology—Technique. 2. Molecular biology—Laboratory manuals. I. Ausubel, Frederick M.

QH506.C87 1987 574.8'8'028 87-21033
ISBN 0-471-50338-X

Printed in the United States of America

20

IMMUNOLOGY

11

Synthetic Peptides for Production of Antibodies that Recognize Intact Proteins

Antibodies that recognize intact proteins can be produced through the use of synthetic peptides based on short stretches of the protein sequence, without first having to isolate the protein. The procedure for selecting stretches of protein sequence likely to be antigenic is relatively straightforward. However, no procedure will identify a single sequence guaranteed to be effective, nor will it usually identify the best single sequence to use. Rather, several sequences will be identified that have a higher-than-average probability of producing an effective antigen.

The steps to produce an effective antibody include: (1) designing the peptide sequence based on the sequence of the protein; (2) synthesizing the peptide; (3) preparing the immunogen either by coupling the synthetic peptide to a carrier protein or through the use of a multiple antigenic peptide (MAP); (4) immunizing the host animal; (5) assaying antibody titer in the host animal's serum; and (6) obtaining the antiserum and/or isolating the antibody. This unit covers steps 1 and 3; step 2 requires a laboratory with expertise in peptide synthesis. Peptide synthesis services are widely available both academically and commercially.

The best method to select potentially effective sequences is via a computer-assisted strategy (see Basic Protocol 1). An alternative manual method is also described (see Alternate Protocol 1) but is not recommended to replace the use of algorithms if there is a choice. A small synthetic peptide is usually insufficiently immunogenic on its own, and two methods have been developed to solve this problem. The first (see Basic Protocol 2) involves chemically coupling the synthetic peptide to a carrier protein to boost the immune response. The second method (see Alternate Protocol 2) entails direct synthesis of a MAP covalent multimer of the simple peptide sequence. Both methods have proven effective and it is a matter of personal preference which to use. Coupling to a carrier protein requires additional chemical manipulations after synthesis of the peptide, while the MAP is complete and ready for immunization at the conclusion of the synthetic protocol. Disadvantages of MAPs are that they are more difficult to produce homogeneously and to analyze postsynthetically. They also may be more prone to insolubility problems.

A carrier protein is a relatively large molecule capable of stimulating an immune response independently. A synthetic peptide coupled to a carrier protein acts as a hapten and produces antibodies specific for the hapten (antibodies against the carrier protein are also produced). The most commonly used carrier proteins are keyhole limpet hemocyanin (KLH) and bovine or rabbit serum albumin (BSA or RSA). KLH is usually preferred, because it tends to elicit a stronger immune response and is evolutionarily more remote from mammalian proteins. A common problem with KLH, however, has been its solubility. Pierce Chemical Company sells a preparation of KLH purported to have better solubility properties (see below).

Alternatively, peptides can be coupled to carrier proteins through either their amino (see Alternate Protocol 3) or carboxyl groups (see Alternate Protocol 4). These two alternate protocols are not recommended as a first choice for coupling, but are included because they have been used successfully and may be advantageous for certain special applications discussed in the Commentary. Also presented are methods for assaying free sulfhydryl content and for reducing disulfide bonds in synthetic peptides (see Support Protocols 1 and 2).

Contributed by Gregory A. Grant
Current Protocols in Molecular Biology (2002) 11.16.1-11.16.19

Once the coupling procedure has been performed, it is possible to determine the approximate degree of coupling by amino acid analysis (see Support Protocol 3). However, in most instances this is unnecessary and the product can be used directly.

BASIC PROTOCOL 1

COMPUTER-ASSISTED SELECTION OF APPROPRIATE ANTIGENIC PEPTIDE SEQUENCES

An antibody produced in response to a simple linear peptide will most likely recognize a linear epitope in a protein. Furthermore, that epitope must be solvent-exposed to be accessible to the antibody. The general features of protein structure that correspond to these criteria are turns or loop structures, which are generally found on the protein surface connecting other elements of secondary structure, and areas of high hydrophilicity, especially those containing charged residues. As a consequence, computer algorithms that predict protein hydrophilicity and tendency to form turns are very useful. Several analytic programs or algorithms that attempt to do this have been developed. Although the choice of method may rely on availability or personal preference, there tends to be a high level of agreement among them. As stated earlier, none of the methods will identify the one single sequence guaranteed to produce an effective antibody against any given protein. Rather, the methods will offer several good candidates, one or several of which can be used.

Many of these algorithms may already be available on a local computer system. They are included in many commercial software packages such as GCG (Genetics Computer Group; see *APPENDIX 4*). The ExPASy Web site of the University of Geneva offers free access to a variety of different programs over the Internet at *http://expasy.org/tools*.

The following protocol utilizes the hydropathy index developed by Kyte and Doolittle (1982) and the secondary structure prediction method for β turns developed by Chou and Fasman (1974) found in the tool "Protscale" at the ExPASy Internet address.

1. Using the selected algorithms, compute the hydropathy index and the tendency for β-turns of the protein sequence. Use a window size of 7 or 9 and give equal weight to each amino acid. Record the results in either graphical or numerical form, or both.

   As an example, the graphical representation of these results for the protein sequence shown in Figure 11.16.1 is presented in Figure 11.16.2.

   A window size determines the number of amino acids to be used in computing a value for the amino acid at the center of the window. For example, a window size of 9 includes 4

```
        10         20         30         40         50         60
MAKVSLEKDK IKFLLVEGVH QKALESLRAA GYTNIEFHKG ALDDEQLKES IRDAHFIGLR
SRTHLTEDVI NAAEKLVAIG CFCIGTNQVD LDAAAKRGIP VFNAPFSNTR SVAELVIGEL 120
LLLLRGVPEA NAKAHRGVWN KLAAGSFEAR GKKLGIIGYG HIGTQLGILA ESLGMYVYFY 180
DIENKLPLGN ATQVQHLSDL LNMSDVVSLH VPENPSTKNM MGAKEISLMK PGSLLINASR 240
GTVVDIPALC DALASKHLAG AAIDVFPTEP ATNSDPFTSP LCEFDNVLLT PHIGGSTQEA 300
QENIGLEVAG KLIKYSDNGS TLSAVNFPEV SLPLHGGRRL MHIHENRPGV LTALNKIFAE 360
QGVNIAAQYL QTSAQMGYVV IDIEADEDVA EKALQAMKAI PGTIRARLLY
```

Figure 11.16.1 The amino acid sequence of a 410-residue protein analyzed by the method presented in Basic Protocol 1. The results are shown in Figure 11.16.2.

amino acids on each side of the central amino acid. The value computed for the central amino acid is the simple average of the values for each amino acid in the window.

2. Compare the results of the two analyses and look for areas of sequence that are high in turn tendency and high in hydrophilicity (low in hydrophobicity).

   In Figure 11.16.2, these areas correspond to positive peaks in the Chou-Fasman analysis and negative peaks in the Kyte-Doolittle analysis. The three best areas in terms of amplitude and correlation are shaded. These correspond to the sequences underlined in Figure 11.16.1. (Note the alignment of these peak optima as compared to the peaks around residue 300.)

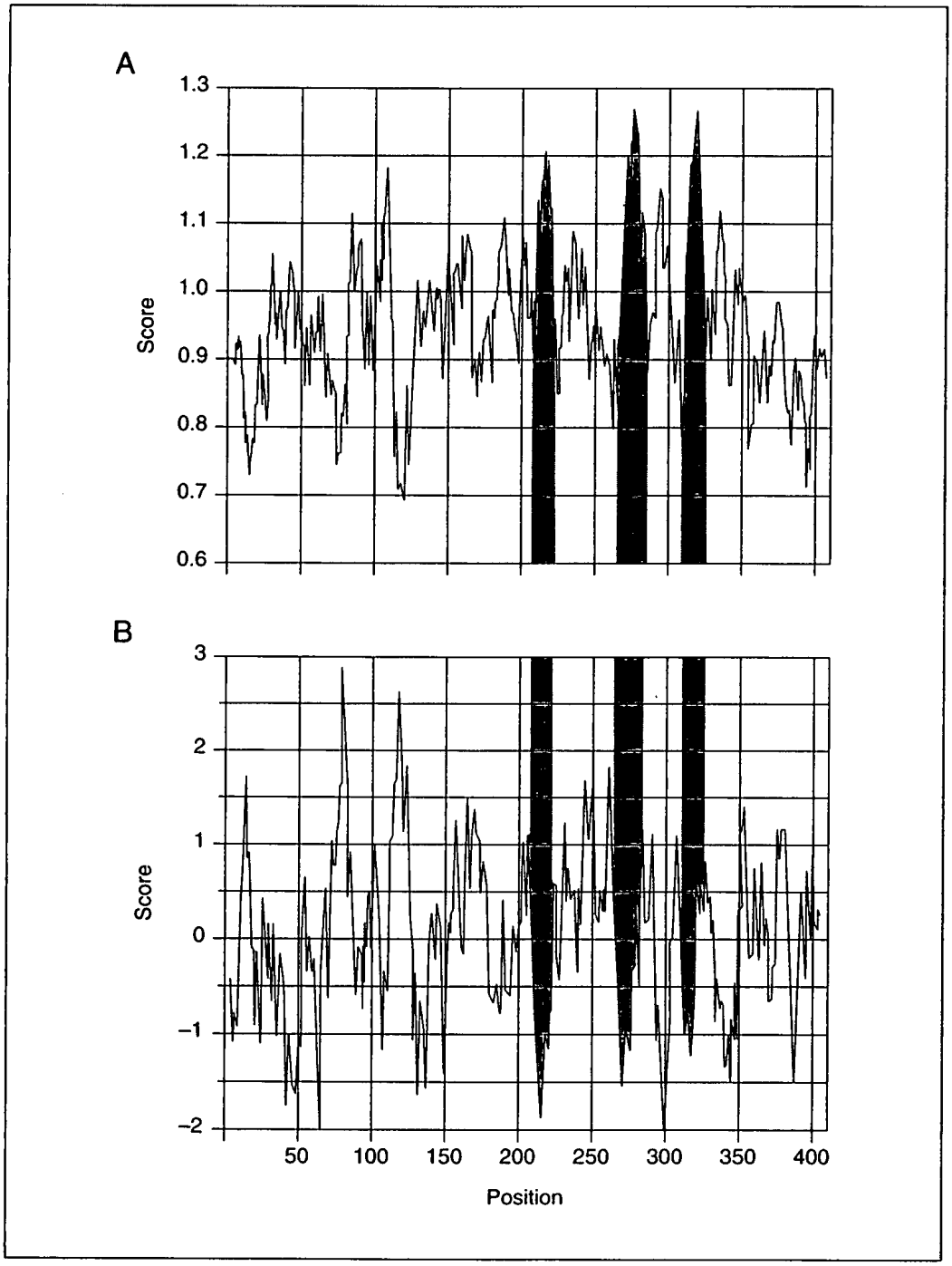


Figure 11.16.2 Graphical representation of the results generated by a computer algorithm for the sequence in Figure 11.16.1, analyzed by the method presented in Basic Protocol 1. The shaded areas represent three regions in the sequence meeting criteria for selection as potential immunogens. (**A**) Analysis for β turns (Chou and Fasman, 1974). (**B**) Analysis for hydrophobicity (Kyte and Doolittle, 1982).

3. Examine the sequences for glycosylation site motifs and discard any sequences that contain them unless it is known that the protein is not glycosylated.

   Amino acids in glycosylated regions may be shielded from presentation to an antibody by masking carbohydrates.

   Amino-linked carbohydrate chains can occur at Asn-X-Ser or Asn-X-Thr sequences. Hydroxyl-linked carbohydrate chains do not appear to have a set motif. A program to assist in the prediction of mucin-type GalNAc O-glycosylation sites in mammalian lipoproteins is found in the tool "NetOGlc" at the Expasy site (http://expasy.org/tools). However, before using read the documentation carefully and keep in mind that such prediction methods cannot always be successful.

4. Select the best sequences resulting from this analysis to use as antigenic peptides. These are sequences where the largest positive values (peaks with positive deflection) for turn propensity correspond in position to the largest negative values (peaks with negative deflection) for hydrophobicity. The values obtained in these analyses are relative and dependent on the individual protein's composition, so it is not possible to set an arbitrary minimum value as a cutoff for rejecting a particular peak. Rather, always select the peaks of greatest magnitude in any given sequence. In addition, the immediate amino-terminal and carboxyl-terminal regions of proteins are often exposed to solvent. If these areas appear to be hydrophilic in nature, they are also acceptable candidates. Thus each analysis may provide several potential sequences. How many peptides to make (see Anticipated Results) is a matter of individual choice.

ALTERNATE PROTOCOL 1

MANUAL INSPECTION TO SELECT APPROPRIATE PEPTIDE SEQUENCES

If computer algorithms are not available, it is possible to select potential sequences by manual inspection. Although there is no evidence that a manual method is any less effective than the use of computer algorithms, there is a greater probability of overlooking potentially important areas of sequence. It is therefore recommended that computer analysis be used whenever it is available. Although it can be done, it would be very time consuming and labor intensive to manually calculate values for every overlapping peptide offset by a single amino acid in the same way that the algorithms do. For this reason, areas rich in polar residues are selected for manual calculation of hydrophilicity and turn propensity.

1. Visually inspect the protein sequence and select areas that contain at least two to three charged residues (Lys, Arg, His, Asp, Glu) within a 10- to 15-residue span.

   If this criterion cannot be met, select sequences with the greatest number of charged residues.

2. From the sequences identified in step 1, select a subset of sequences that are the highest in Ser, Thr, Asn, Gln, Pro, and Tyr content.

3. Calculate average hydrophilicity and turn propensity for each amino acid in the selected sequences using the values given in Table 11.16.1 and a window of 9 residues (see Basic Protocol 1, step 1).

   Be sure to include the residues flanking the selected sequence for calculation of values for the residues at the ends of the selected sequence. In other words, do not use different size windows.

4. Plot the values for each amino acid of a chosen sequence.

   Sequences whose optimal values for hydrophilicity and turn propensity correspond (as in Fig. 11.16.2) are considered good candidates.

5. Inspect sequences for glycosylation motifs and discard these candidates (see Basic Protocol 1, step 3).

Table 11.16.1 Hydrophobic and β-Turn Indices of Amino Acids

Amino acid	Symbols	Hydrophobicity value[a]	β-turn propensity[b]
Arginine	Arg (R)	–4.5	0.95
Lysine	Lys (K)	–3.9	1.01
Aspartic acid	Asp (D)	–3.5	1.46
Glutamic acid	Glu (E)	–3.5	0.74
Asparagine	Asn (N)	–3.5	1.56
Glutamine	Gln (Q)	–3.5	0.98
Histidine	His (H)	–3.2	0.95
Proline	Pro (P)	–1.6	1.52
Tyrosine	Tyr (Y)	–1.3	1.14
Tryptophan	Trp (W)	–0.9	0.96
Serine	Ser (S)	–0.8	1.43
Threonine	Thr (T)	–0.7	0.96
Glycine	Gly (G)	–0.4	1.56
Alanine	Ala (A)	1.8	0.66
Methionine	Met (M)	1.9	0.60
Cysteine	Cys (C)	2.5	1.19
Phenylalanine	Phe (F)	2.8	0.60
Leucine	Leu (L)	3.8	0.59
Valine	Val (V)	4.2	0.50
Isoleucine	Ile (I)	4.5	0.47

[a]Kyte and Doolittle (1982).

[b]Chou and Fasman (1974).

Amino acids of glycosylated regions may be masked in native proteins, so an antibody raised against them would be ineffective.

6. Select the best sequences (see Basic Protocol 1, step 4 for criteria), choosing a high turn-propensity-to-hydrophobicity ratio.

DESIGNING A SYNTHETIC PEPTIDE FOR COUPLING TO A CARRIER PROTEIN

BASIC PROTOCOL 2

Although there is no direct evidence to show that the state of the termini of the peptide affects its ability to produce antibodies that will react with the protein, most procedures suggest that the termini of the peptide should mimic their native state. Thus, sequences whose terminal residues normally are in peptide linkage in the protein can have their amino-terminal and carboxyl-terminal groups modified by acetylation and amidation, respectively, during synthesis.

Modification of the amino or carboxyl termini will decrease the polarity of the peptide in solution and could have a significant effect on the peptide's solubility. If the peptide lacks sufficient protonatable side chains, modification of the termini can be omitted. A general rule to predict solubility is that the total number of charges at a given pH should be at least 20% of the number of residues in the peptide.

1. Choose a sequence of 10 to 15 amino acid residues for the synthetic peptide.

   Longer peptides are more difficult and expensive to make, and they are usually unnecessary.

REAGENTS AND SOLUTIONS

Use Milli-Q-purified water or equivalent for the preparation of all buffers. For common stock solutions, see APPENDIX 2; for suppliers, see APPENDIX 4.

Cysteine standard stock solution

Dissolve 26.3 mg cysteine hydrochloride monohydrate in 100 ml of 0.1 M sodium phosphate, pH 8.0 (*APPENDIX 2*). Prepare immediately before use.

Ellman's reagent solution

Dissolve 4 mg Ellman's reagent, 5,5′-dithio-bis-(2-nitrobenzoic acid) (Pierce), in 1 ml of 0.1 M sodium phosphate, pH 8.0 (*APPENDIX 2*). Prepare immediately before use.

Glutaraldehyde solution, 0.15%

Add 30 µl of 25% aqueous glutaraldehyde solution to 5 ml of 50 mM sodium borate buffer, pH 8.0 (pH adjusted with HCl). Prepare fresh and use immediately. If the glutaraldehyde precipitates, check the pH. It should not be above 8.0; a slightly lower pH can be used (pH 7 to 8).

CAUTION: *Glutaraldehyde is a sensitizing agent that should be handled in a hood and only according to the recommendations in the Material Safety Data Sheet. When mixing solutions or performing reactions, keep the container covered to prevent vapors from escaping into the atmosphere.*

Guanidine·HCl, 6 M

Dissolve 1 g guanidine·HCl in 1 ml of 0.05 M sodium phosphate, pH 7.0 (*APPENDIX 2*). Store up to several weeks at room temperature.

The resulting 1.8-ml solution should be ~0.025 M phosphate/6 M guanidine·HCl at pH 7.0.

COMMENTARY

Background Information

Synthetic peptides are linear arrays of amino acids that in most instances possess a random structure in solution. While it is not difficult to produce antipeptide antibodies, it does not necessarily follow that the antibodies will recognize a protein containing the same stretch of sequence found in the peptide. In order for this to occur, the amino acids in the protein must be oriented to the antibody in a way similar to that of the synthetic peptide. This generally requires three basic features of the protein: (1) that the stretch of sequence be exposed to solvent; (2) that the sequence be a continuous stretch of amino acids; and (3) that it not possess a higher-order structure that renders it unrecognizable by the antibody population.

The large number of model protein structures now available indicate that almost all of the ionized groups in water-soluble proteins are on the protein surface. Asp, Glu, Lys, and Arg residues, on the average, comprise 27% of the protein surface and only ~4% of the protein interior. The fraction of residues that are at least 95% buried range from 0.36 to 0.60 for nonpolar residues and 0.01 to 0.23 for polar residues. Only 1% of Arg and 3% of Lys residues fall into the 95% buried range (Creighton, 1993). Therefore, it is reasonable to expect solvent-exposed areas of proteins to display relatively high levels of polar and charged residues, particularly Arg and Lys.

Proteins display three kinds of secondary structure: α-helices, β-sheets, and turns or loops. Turns or loops generally connect elements of α-helices and β-sheets, and can either fit one of several rather strict motifs with recognizable hydrogen bonding patterns or be of a more extended, random nature. These turn or loop structures appear to be most useful for antibody production because they tend to be found on the surface of proteins connecting larger arrays of helices and sheets, and they consist of continuous stretches of amino acids. Although many amino acid residues in helices and sheets are also exposed at the surface, the regular geometry of amino acids contained within them makes them less suitable for this purpose. For instance, in β-sheet structures the side chain of each successive amino acid in the β-sheet strand points in the opposite direction to the ones immediately preceding and following it. Thus, even if the amino acid side chains are not predominantly buried in the interior of

the protein, only every other side chain is exposed on the same surface of the sheet. This can hinder recognition by an antibody produced with a linear peptide capable of assuming a more random structure. A similar situation exists for α-helices. Although the change in direction of the side chains of successive amino acids is perhaps not as abrupt as in β-sheets, only approximately every third or fourth side chain is found on the same surface of the helix. Epitopes in proteins have been identified in amphipathic helices, but unless the synthetic peptide assumes a similar helical structure in solution, recognition by the antibody may be problematic.

These considerations have led to more useful methods for predicting sequences that will produce antibodies recognizing intact proteins. A variety of different indices that predict hydrophilicity or hydrophobicity and secondary structure are available. In addition, predictive methods based on segmental mobility, side chain accessibility, and sequence variability (see Van Regenmortel et al., 1988) have also been proposed. All of these methods generally tend to yield similar results, but it must be noted that these procedures were developed for (and work best with) water-soluble proteins composed of a single globular structure. Additional complications can arise with multisubunit proteins, where normally exposed structures may be shielded by subunit interactions, or membrane proteins with large sections shielded from the solvent.

The method presented in this unit utilizes the correlation between the hydrophilic character of a peptide sequence (Kyte and Doolittle, 1982) and its propensity to form β-turn structures (Chou and Fasman, 1974). Free access to these and many other algorithms is provided at the ExPASy Web site of the University of Geneva at *http://expasy.org.tools*.

After selection of the peptide sequence, an effective immunogen is generally produced by coupling the peptide to a carrier protein or by synthesizing a multiple antigenic peptide (MAP), with four or eight identical peptides assembled simultaneously on the α and ε amines of the terminal lysines of a branched core (see Fig. 11.16.3).

Critical Parameters

Analyzing protein sequences with algorithms or tables of assigned values for amino acids is a well-established procedure, but evaluating these results and selecting the candidate sequences requires some consideration. To take full advantage of the results, choose areas of sequence that give the maximum values for the properties being evaluated and that also show the highest degree of residue-by-residue correlation. In other words, choose areas of maximum amplitude where the centers of the peaks correspond to the same sequence with a divergence of no more than two to three residues. Examples of this are given in Figure 11.16.2, which shows results from the method presented in Basic Protocol 1 for the sequence shown in Figure 11.16.1. The top panel in Figure 11.16.2 predicts β-turns as calculated by the method of Chou and Fasman (1974). The bottom panel is a prediction of hydrophobicity using the parameters of Kyte and Doolittle (1982). The data are analyzed by looking for areas of high turn propensity (maximum positive deflection in the top panel) and high hydrophilicity (maximum negative deflection in the bottom panel). The shaded areas in Figure 11.16.2 designate three segments that meet these criteria. Note that the maximum and minimum values of these three stretches of protein sequence correlate very well. Additional areas of high hydrophilicity (bottom panel) are found near residues 64, 132, 137, 149, and 345, although the β-turn values of these secondary candidates are not as high as those of the three shaded areas. Two equally hydrophilic areas at residues 49 and 299 correspond to downward deflections in the β-turn profile and are thus not good candidates based on this analysis.

Many different chemistries are available for coupling synthetic peptides to carrier proteins to produce effective immunogens (Van Regenmortel et al., 1988). In many cases, however, side reactions or incompatibilities in chemistry between the coupling agent and the residues present in the peptide can be problematic. In order to simplify the process and present the greatest probability of success in most cases, only a few coupling methods are presented in this unit. In this regard, the recommended coupling procedure is cross-linking of the peptide via cysteine residues to keyhole limpet hemocyanin (KLH) with the heterobifunctional reagent *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS; see Basic Protocol 3). This effective method has enjoyed great success and can be used for virtually any peptide. The one caveat is that it is not recommended for peptides with internal cysteine residues, since they will also link to the carrier. It is also critical, when coupling with MBS through an added terminal

cysteine residue, that the sulfhydryl group of the peptide be present in the free or reduced form (see Support Protocols 1 and 2).

In addition to MBS coupling, other procedures commonly used (see Alternate Protocols 3 and 4) are included as alternatives for use in special situations, but these are not recommended as a general alternative to MBS because they are more restrictive and have the potential for undesirable side reactions. Glutaraldehyde coupling (see Alternate Protocol 3) should not be used with peptides containing internal Lys, Cys, Tyr, or His residues and, since it is a homobifunctional reagent, cross-linking of the peptide to itself and the carrier to itself can occur. The latter lowers antigenicity and can result in extensive aggregation and precipitation of the carrier. 1-ethyl-3(3-dimethylaminopropyl) carbodiimide (EDC; see Alternate Protocol 4) is a water-soluble carbodiimide and should not be used with peptides containing internal Lys, Glu, Asp, Tyr, or Cys residues. Alternate Protocol 5 describes a simple photochemical coupling strategy (Gorka et al., 1989).

Another good alternative for most peptides is the production of a multiple antigenic peptide (MAP; see Alternate Protocol 2). With this method the composition of the peptide is not a concern beyond its potential solubility properties. In most cases, since hydrophilic sequences are selected, this also is not a major problem. Both four- and eight-branched MAPs have been found to be effective. However, four-branched MAPs are recommended because they are less prone to synthesis problems and are easier to characterize.

As with any synthetic peptide, the product must be well characterized before use. If the peptide is not what it was intended to be, this decreases the probability of generating antibodies that will recognize the protein. At the very least, check synthetic peptides for homogeneity by analytical HPLC and correct mass by mass spectrometry (see *UNITS 10.21 & 10.22*). Characterization of MAP can be more problematic due to their multibranched nature (Mints et al., 1997): HPLC and mass spectrometric analysis can be compromised by the presence of four to eight peptide chains per molecule, each of which may have only a small percentage of modification at any particular residue but which in the aggregate contribute to broad spectra. However, this feature of MAPs usually does not tend to compromise their ability to form antigens of the proper peptide since the correct sequence is usually present in high enough concentration that a significant amount of specific antibody is produced among the polyclonal population. Amino acid analysis (*UNIT 10.1B*), which is less sensitive to multiple small differences, tends to give a reasonable assessment of the MAP integrity.

Anticipated Results

The methods outlined in this unit produce an effective polyclonal antiserum against an intact protein from a single peptide sequence ~50% to 70% of the time. Therefore, it is advisable to prepare two or three different peptides from a given protein to increase the probability of at least one of them being effective.

Time Considerations

Computer-assisted analysis of a protein sequence and inspection of the data to select several candidate sequences takes from 5 to 30 min. Manual analysis of a protein sequence can take several hours but can certainly be accomplished in <1 day. Selection of peptide design and manner of synthesis as well as selection of a coupling method will take <1 hr. Actual preparation of the peptide can be accomplished in 3 to 4 days, but this may vary depending on the turnaround time of the synthetic laboratory. Coupling a synthetic peptide to a carrier protein takes from 1 to 2 days. Although not covered in this unit, production of the antisera will vary with the animal and protocol used, but generally requires 2 to 3 months. It is therefore advisable to inject several animals with different peptides at one time.

Literature Cited

Chou, P.Y. and Fasman, G.D. 1974. Prediction of protein conformation. *Biochemistry* 13:222-245.

Creighton, T.E. 1993. Proteins: Structure and Molecular Properties, 2nd ed. W.H. Freeman, New York.

Gailit, J. 1993. Restoring free sulfhydryl groups in synthetic peptides. *Anal. Biochem.* 214:334-335.

Getz, E.B., Xiao, M. Chakrabarty, T., Cooke, R., and Selvin, P.R. 1999. A comparison between the sulfhydryl reductants Tris(2-carboxyethyl) phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.* 272:73-80.

Gorka, J., McCourt, D.W., and Schwartz, B.D. 1989. Automated synthesis of a C-terminal photoprobe using combined Fmoc and *t*-Boc synthesis strategies on a single automated peptide synthesizer. *Peptide Res.* 2:376-380.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.

Mints, L., Hogue Angeletti, R., and Nieves, E. 1997. Analysis of MAPS peptides. *ABRF News* 8:22-26.

Posnett, D.N., McGrath, H., and Tam, J.P. 1988. A novel method for producing anti-peptide antibodies. Production of site-specific antibodies to the T-cell antigen beta-chain. *J. Biol. Chem.* 263:1719-1725.

Tam, J.P. 1988. Synthetic peptide vaccine design: Synthesis and properties of a high density multiple antigenic peptide system. *Proc. Natl. Acad. Sci. U.S.A.* 85:5409-5413.

Van Regenmortel, M.H.V., Briand, J.P., Muller, S., and Plaué, S. 1988. Synthetic polypeptides as antigens. *In* Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 19 (R.H. Burdon and P.H. van Knippenberg, eds.). Elsevier/North-Holland, Amsterdam.

Key References

Van Regenmortel et al., 1988. See above.

Comprehensive treatment of theory and method.

Tam, J.P. 1988. High density multiple antigen peptide system for preparation of anti-peptide antibodies. *Methods Enzymol.* 168:7-15.

Original methods article for MAPs.

Internet Resource

http://expasy.org/tools.html

Web site for programs to analyze protein sequences.

Contributed by Gregory A. Grant
Washington University School of Medicine
St. Louis, Missouri